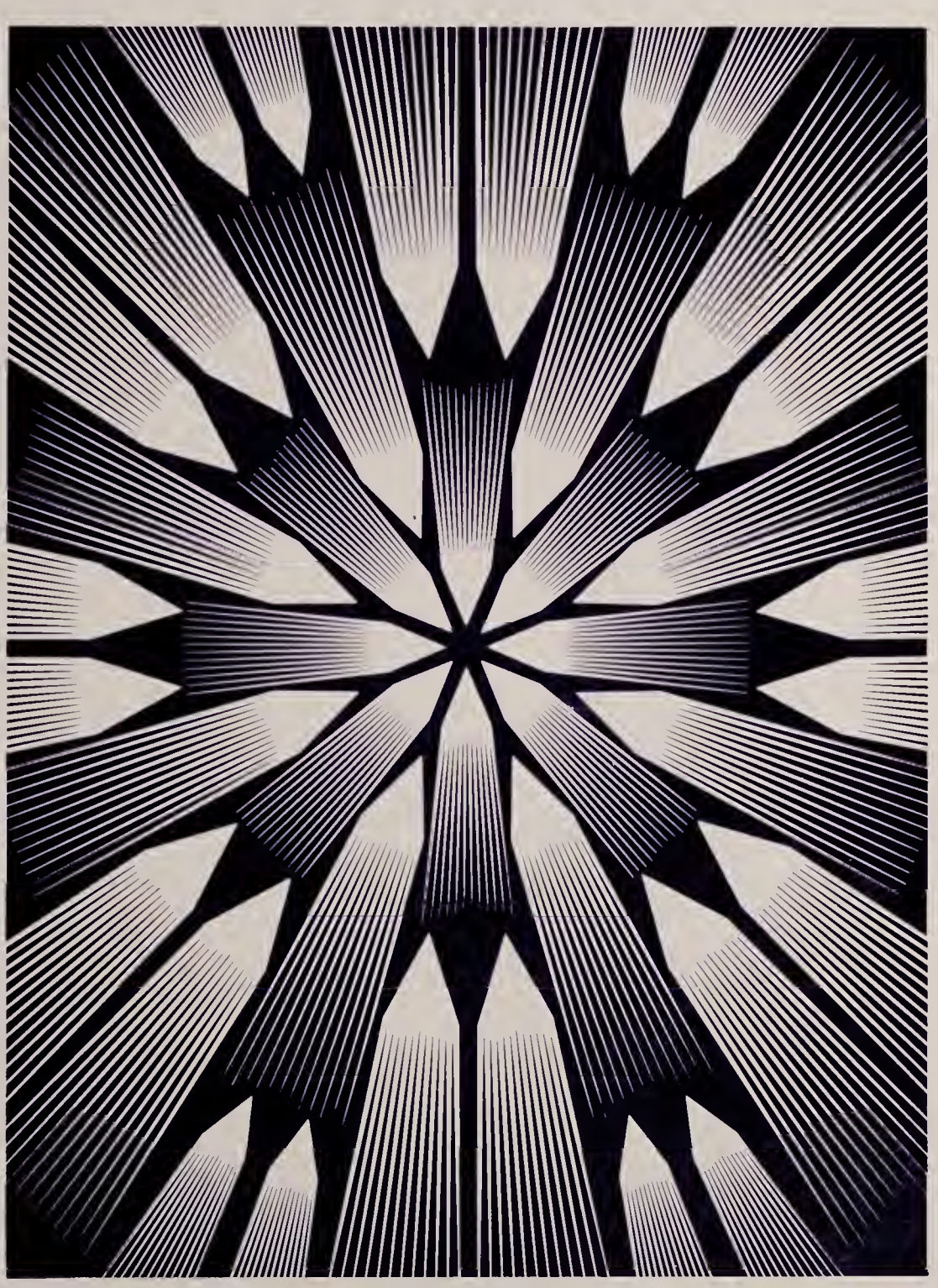
# Measuring the Quality of Life of People With Visual Impairment

## Proceedings of a Workshop



HV1714
D844
C. 2

U.S. Department of Health and Human Services
Public Health Service ■ National Institutes of Health

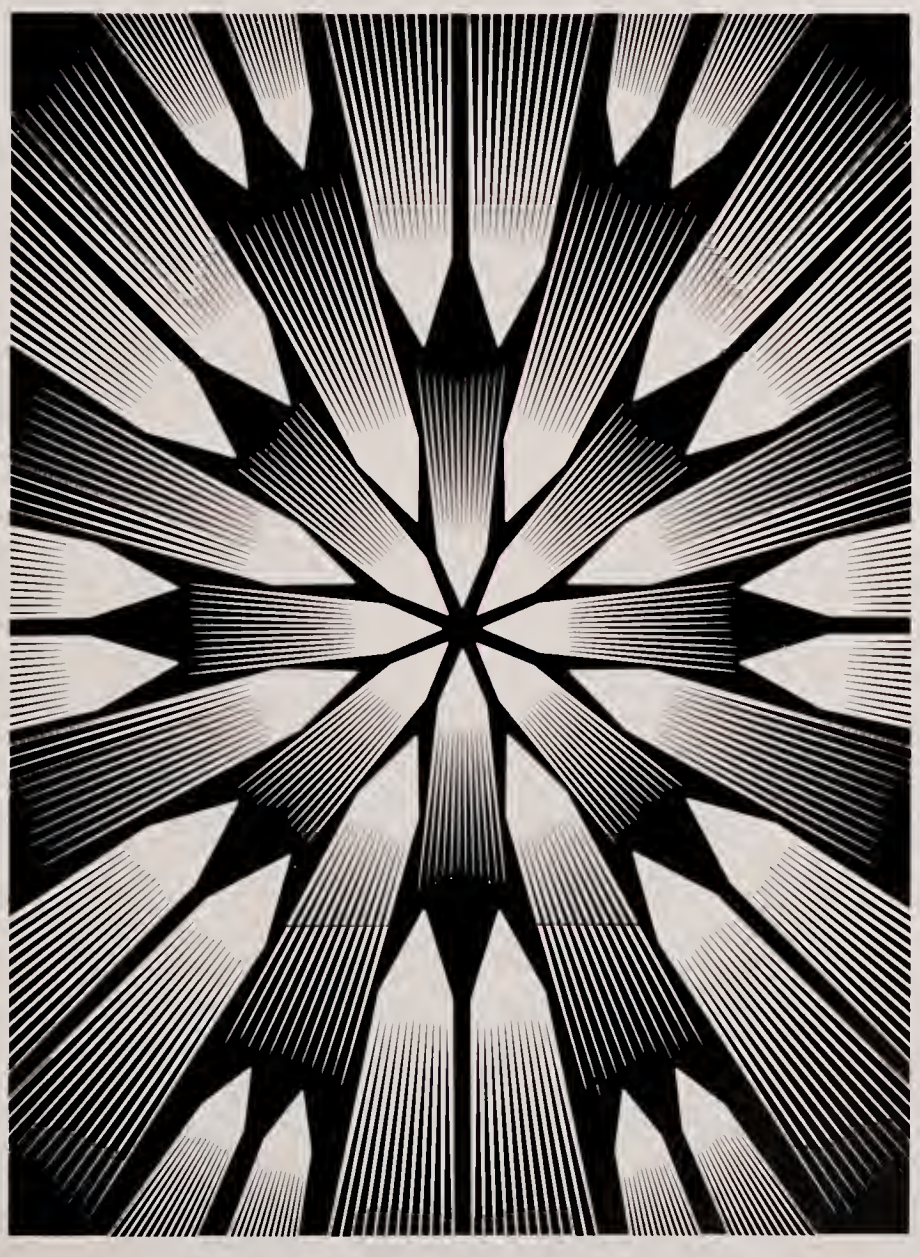
EX
LIBRIS
M.C. MIGEL LIBRARY
AMERICAN PRINTING
HOUSE FOR THE BLIND

C. 2

# Measuring the Quality of Life of People With Visual Impairment

## Proceedings of a Workshop

Michael F. Drummond, Ph.D.
Editor



U.S. Department of Health and Human Services
Public Health Service ■ National Institutes of Health
National Eye Institute
NIH Publication No. 90-3078
October 1990

HV1714
D844
C. 2

# Contents


| | | |
|---|---|---|
| iii | **Participants** | |
| 1 | **Executive Summary** | |
| 7 | **Introduction to the Workshop** | *Carl Kupfer, M.D.* |
| | **Papers** | |
| 9 | Measuring and Valuing States of Health Relating to Visual Impairment: Background and Overview | *Martin Buxton* |
| 13 | How Visual Impairment Affects Quality of Life: Experiences with Quality-of-Life Measurements in Clinical Ophthalmology<br>*Discussion* | *Peter Bernth-Petersen, M.D., Ph.D.* |
| 17 | How Visual Impairment Affects Quality of Life<br>*Discussion* | *Gordon E. Legge, Ph.D.* |
| 23 | Measures of Outcome in Current Clinical Trials of Eye Care<br>*Discussion* | *Ronald Klein, M.D., M.P.H.* |
| 29 | General Health Measurement and Quality of Life in Visual Impairment<br>*Discussion* | *Ian McDowell, Ph.D.* |
| 37 | Disease-Specific Versus General Health State Measures<br>*Discussion* | *Donald L. Patrick, Ph.D.* |
| 51 | Incorporating Health Status Measures into Clinical Trials<br>*Discussion* | *John E. Ware, Jr., Ph.D.* |
| 59 | The Utility Approach to Measuring Health States<br>*Discussion* | *Michael F. Drummond, Ph.D.* |
| 65 | Major Themes and Conclusions: A Strategy for Measurement of Quality of Life in National Eye Institute Trials | *Michael F. Drummond, Ph.D.*, and *Frederick L. Ferris, III, M.D.* |
| | **Appendices** | |
| 69 | A. Measuring and Valuing Health States Relating to Visual Impairment: A Review of Literature, Concepts, and Methods | *Brian Ferguson, Martin Buxton,* and *Michael F. Drummond, Ph.D.* |
| 83 | B. Selected Annotated Bibliography of Clinical Studies and Surveys of the Effects of Eye Disease or Its Treatment | |

# Participants

**Co-Chairmen**

*Michael F. Drummond, Ph.D.*
Director, Health Services Management Centre
University of Birmingham
Birmingham
United Kingdom

*Frederick L. Ferris, III, M.D.*
Clinical Trials Branch
Biometry and Epidemiology Program
National Eye Institute
Bethesda, MD

---

*Constance W. Atwell, Ph.D.*
Extramural and Collaborative Programs
National Eye Institute
Bethesda, MD

*Ivan Barofsky, Ph.D.*
Francis Scott Key Medical Center
The Johns Hopkins School of Medicine
Baltimore, MD

*Marilyn Bergner, Ph.D.*
Health Services Research Center
The Johns Hopkins University School of Hygiene and Public Health
Baltimore, MD

*Peter Bernth-Petersen, M.D., Ph.D.*
University Eye Clinic
Rigshospitalet
Copenhagen
Denmark

*Martin Buxton*
Health Economics Research Group
Brunel University
Uxbridge, Middlesex
United Kingdom

*Connie L. Carmack, M.S.*
Office of Planning and Reporting
National Eye Institute
Bethesda, MD

*Emily Chew, M.D.*
Biometry and Epidemiology Program
National Eye Institute
Bethesda, MD

*Michael P. Davis, M.S.*
Office of Planning and Reporting
National Eye Institute
Bethesda, MD

*Coralie Farlee, Ph.D.*
Fogarty International Center
National Institutes of Health
Bethesda, MD

*Dennis G. Fryback, Ph.D.*
Department of Industrial Engineering
University of Wisconsin
Madison, WI

*Terrence R. Gillen, M.A., M.B.A.*
Office of Planning and Reporting
National Eye Institute
Bethesda, MD

*Barbara S. Hawkins, M.S.*
COMS Coordinating Center
Wilmer Eye Institute
The Johns Hopkins School of Medicine
Baltimore, MD

*Jonathan C. Javitt, Ph.D.*
Dana Center for Preventive Ophthalmology
Wilmer Eye Institute
The Johns Hopkins School of Medicine
Baltimore, MD

*Ronald Klein, M.D., M.P.H.*
Department of Ophthalmology
University of Wisconsin
Clinical Science Center
Madison, WI

*Carl Kupfer, M.D.*
National Eye Institute
Bethesda, MD

*Gordon E. Legge, Ph.D.*
Department of Psychology
University of Minnesota
Minneapolis, MN

*Kathleen N. Lohr, Ph.D.*
Institute of Medicine
National Academy of Sciences
Washington, DC

*Ian McDowell, Ph.D.*
Department of Community Medicine
University of Ottawa
Ottawa, Ontario,
Canada

*Jack A. McLaughlin, Ph.D.*
Extramural and Collaborative Programs
National Eye Institute
Bethesda, MD

*Edward H. McManus, M.P.A.*
Office of the Director
National Eye Institute
Bethesda, MD

*Julian M. Morris*
Office of Planning and Reporting
National Eye Institute
Bethesda, MD

*Robert P. Murphy, M.D.*
Wilmer Eye Institute
The Johns Hopkins School of Medicine
Baltimore, MD

*Donald L. Patrick, Ph.D.*
Department of Health Services
University of Washington
Seattle, WA

*Lawrence I. Rand, M.D.*
Elliot P. Joslin Research Laboratory
Joslin Diabetes Center
Boston, MA

*William E. Rhode, Ph.D.*
Office of Science Policy and Legislation
Office of the Director
National Institutes of Health
Bethesda, MD

*James Schuttinga, Ph.D.*
Office of Science Policy and Legislation
Office of the Director
National Institutes of Health
Bethesda, MD

*John E. Ware, Jr., Ph.D.*
Institute for the Advancement of Health and Medical Care
New England Medical Center
Newton, MA

## Acknowledgments

I would like to express my gratitude to my co-chairman, Dr. Rick Ferris, and all the participants in this workshop, especially those who presented papers and took part in the discussion. Special thanks are due Mr. Michael P. Davis of the National Eye Institute staff for his outstanding meeting arrangements and logistical support, and Mr. Julian M. Morris of the NEI, Ms. Nancy Klein of CSR, Incorporated and Mr. James B. Lawrence of University Research Corporation for their assistance in the editorial preparation and graphic design of these proceedings. Finally, I would like to thank Dr. Carl Kupfer and Mr. Edward McManus, NEI Director and Deputy Director respectively, for their foresight in suggesting that the workshop be held, for consulting on the topics to be covered, and last, but not least, for providing the meeting site and funding.

Michael F. Drummond, Ph.D.
Editor

# Executive Summary

Interest in quality-of-life measurement is growing for two major reasons. First is the increasing recognition within the scientific medical community that the most meaningful outcome of a clinical trial may be the patient's own perception of changes in his or her well-being. Demonstrating a way to improve the quality of life may be a key benefit of a trial, being just as important as such traditional outcome measures as changes in clinical manifestations and increased longevity. Second, a growing body of excellent information on health status and quality-of-life measurement is now available for assessing such outcomes.

Policymakers and decisionmakers are particularly interested in obtaining evidence on the relative values of different medical interventions. The National Eye Institute (NEI), one of the National Institutes of Health, has conducted analyses of the economic consequences of its research programs for a number of years. In recent years the Institute has emphasized that the value of treatments for eye disease should not be measured only in monetary terms. Measures that can help compare the impact of medical interventions on the quality of life are also important both for patients and planners.

The NEI sponsored a workshop on "Measuring the Quality of Life of People with Visual Impairment" on September 13-14, 1988. The workshop had two purposes: 1) to review current issues and knowledge in the field of quality-of-life assessment, with special reference to how they may apply to conditions of the visual system; and 2) to evaluate existing scales for their potential usefulness in determining the effect on the quality of life of new treatments studied in clinical trials supported by the NEI.

## Application of Quality-of-Life Measurements in Government Decision-Making

As reported at the workshop, quality-of-life measurements are being used by government decisionmakers in various countries. For example, the government of Denmark has increased its allocation of resources to cataract surgery in part as a result of the use of the Visual Functioning Index. This instrument was designed to measure the benefit to patients of cataract surgery, including its impact on the occupational, social, and economic aspects of quality of life. Findings demonstrated the advantages of cataract surgery at an earlier stage in cataract development than was usual in Denmark.

## Measurement of Health Status In Relation to Visual Function

The relationship between health-related quality of life and visual function varies according to duration of impairment, age, and a range of socioeconomic and psychological variables. In the assessment of the relationship, both objective and subjective measures may be used. Whether one type of measure is more meaningful than the other in assessing quality of life is still a subject of debate, but clearly both are necessary components of quality-of-life measurements.

Well-validated measures of health status could be valuable in a range of situations, such as clinical trials, trials specifically designed to measure quality of life, cost-effectiveness studies, multipurpose experimental studies, and health care evaluation and monitoring. In all applications, the purpose of obtaining quality-of-life

information should be considered. Quality-of-life measures can be used in two ways: to measure change or difference in outcomes *within* a trial, for which purpose, discriminatory power is a key requirement; or to relate to data on other health care interventions, in which case common measurement and comparability are the main requirements.

## How Visual Impairment Affects Quality of Life

Visual impairment causes difficulties with everyday life, such as reading, writing, driving, general mobility, and watching television. Other daily problems may include social awkwardness and coping with the stigma—actual or perceived—of low vision or blindness. Visual impairment also can affect quality of life in long-term ways, closing some career and educational options, and limiting social life. Some of these factors can be documented by objective measures although concerns such as stigma and social difficulties may require a more subjective analysis.

## Traditional Outcome Measures

Traditional measures used to assess the outcome of treatment for eye disease include changes in clinical manifestations of disease and measures of visual function. Tests of visual acuity are among the most common outcome measures in current vision-related clinical trials. Also important are measures of changes in visual field, contrast sensitivity, and color vision.

## Categories of Quality-of-Life Measurement

Instruments that assess health-related quality of life include measures incorporating six major concepts: opportunities, perceptions and life satisfaction, social function, psychological function, physical function, and impairment. The instruments fall into three main categories:

- *Functional status measures.* This category includes the familiar activities of daily living (ADL) scales, which are very often brief questionnaires covering commonly required activities such as climbing stairs, washing, and so forth. The vast majority have not been evaluated very closely for reliability or validity.
- *General health status measures.* Unlike functional status indicators, general health status measurements take a multi-dimensional approach. They include the familiar ADL themes that characterize the functional status measurements but also usually cover psychological, physical, and often social functioning. In addition, they sometimes cover some quality-of-life dimensions such as the extent to which social support is available and the resources that the environment can offer the patient.

  General health status measures are generally more recent and better validated than functional status measures, and use more precise scoring procedures. They tend to be more suited to research applications than to brief clinical assessments.
- *Quality-of-life measures.* These scales cover the topics covered in the two categories described above, but also include personal, material, and environmental resources and opportunities for meeting needs. Generally defined as an overall evaluation of the subjective experience of life, quality-of-life scales are broader than general health measurements. The main contrast between health status and quality-of-life measurements lies in the attention paid to environmental factors in the latter.

Instruments range from very short (18 items) to very long (more than 300 items); may be designed for self-administration or administration by an interviewer; provide ordinal, interval, or ratio scales; and may be intended for use in clinical settings, surveys, and other studies.

There are three main groups of general instruments:

- *Health profiles,* which incorporate a number of different dimensions of health, e.g., energy, social interaction, mobility, or emotional reactions, and provide an overall picture of health.
- *Health indices,* which bring these various dimensions together and express health status as a single point on an index.

- *Utility measures,* which attempt to show the relative value of different points on a single index, using an interval or ratio scale. Of these, utility measures are of particular interest to health economists and policymakers.

### *Disease-Specific Versus Generic Measures*

Instruments that measure quality of life also can be categorized as disease-specific or generic measures. Generic health status measures are those that are broadly applicable to various types and severities of disease, medical treatments or health interventions, and demographic and cultural subgroups. The Sickness Impact Profile, for example, has been used with patients having a wide range of conditions, including visual impairment.

Disease-specific measures are those designed to assess specific diagnostic groups or patient populations, often with the goal of measuring responsiveness or clinically important changes.

Of the disease-specific instruments, few address visual impairment. Many, however, focus on general functional disability, including mobility and may prove sensitive to levels of visual impairment. Disease-specific measures tend to use wording and instructions tailored to the diagnostic group, condition, or population under investigation. Items in arthritis-specific measures, for example, may include the phrase "due to my arthritis."

### *Models for Using Generic and Disease-Specific Measures*

Four different models seem relevant in examining the use of generic and disease-specific measures to date.

- Include both generic and disease-specific measures in the same investigation even though the concepts covered by the different instruments may overlap substantially.
- Compare a generic instrument and a generic instrument modified for the specific population of interest.
- Use a generic health status instrument with a condition-specific supplement with little overlap.
- Administer health status batteries, which are collections of specific measures that are scored independently.

### *Utility Measures*

Utility measures are of particular interest to health economists and policymakers. The utility approach differs from most other approaches to measuring health-related quality of life. This approach yields a single value for health states on an interval scale. On such a scale, death equals zero and complete well-being equals one. When combined with data on the years of life gained by health care interventions, the utility values for health states enable the calculation of the number of *quality-adjusted life-years* (QALYs) gained from therapy.

This generic measure of the output of health care programs has gained popularity with health economists because they can use it to compare interventions in terms of their incremental cost per QALY gained. The proponents of health utility measures contend that they are most useful for making planning decisions, rather than individual treatment decisions.

Three main methods of obtaining utility values are the rating scale, the time tradeoff, and the standard gamble. Each method generates values by asking the respondent, who may be a patient, doctor, or member of the general public, to compare health states that are described to him or her.

In using utility values, the analyst should consider *whose* values are most relevant. For planning decisions one might argue that the values of the general public are the most relevant. On the other hand, the general public might be more interested in how patients value different health states. Experiencing a particular illness, after all, is different from responding to hypothetical scenarios.

## Selection Criteria

Both generic and disease-specific assessments are useful for clinical trials. The choice of measures depends on the objectives of measurement and environment of the application. In designing an optimal strategy, investigators need

to identify and assign priorities to measurement objectives.

Generic measures permit comparison of different populations and different programs, an important objective for policy analysis and decision-making. By contrast, instruments tailored to specific diseases, conditions, and populations are critical for identifying important concerns of patients and for measuring certain, clinically important changes resulting from specific treatments. Experience with disease-specific measures to date indicates their usefulness in discriminating among different conditions and in assessing changes.

Measures may be used to discriminate among patients at a point in time, to predict future outcomes or events, and to measure changes over time. Technical requirements include:

- *Validity.* Generic measures used in condition- or population-specific applications may have low content validity because they contain items of little or no relevance to study participants. By contrast, the focus of disease-specific measures gives them high content validity for clinical investigations of special populations. The construct validity of disease-specific measures is less well established than that of generic instruments.
- *Reliability.* Global measures created from generic health status measures often have high test-retest and internal consistency reliability. Studies comparing the test-retest reliability of disease-specific measures with that of generic measures have not yet been published.
- *Responsiveness.* Disease-specific measures with items selected to assess particular concerns or worded to attribute change in the condition of interest, e.g., "due to my vision," may be particularly sensitive to within-subject changes and thus more responsive than generic measures that contain items unrelated to change.
- *Ease of analysis.* Changes in generic health measures are more difficult to interpret than changes in disease-specific measures, although even small changes in portions of such measures may be quite useful; for example, changes in physical mobility or self-care are meaningful in disabled populations. Changes in scores on the most general measures, such as health perceptions or global physical and psychosocial dimension scores, can be even more difficult to interpret.
- *Generalizability.* Generic measures provide a denominator or common unit of outcome by which to judge the relative severity of health outcomes and the relative effectiveness of health interventions. Disease-specific measures developed for special populations are not intended for general application. The issue of generalizability may still arise, however, in relation to the disease, condition, or population under investigation. Arthritis-specific measures have been applied across a large number of investigations. The cross-cultural content of health status measures is relevant to both types of measures.

Practical considerations, such as cost, time, and the burden on the respondent may also be factors in selecting an instrument.

## Major Issues and Recommendations

The major workshop conclusion was that measurement of health-related quality of life is a complex task that must be approached with due caution. A five-point strategy was offered for measurement of quality of life in National Eye Institute-supported clinical trials:

1. Quality-of-life measures should not be viewed as a replacement for traditional clinical measures of treatment success, such as visual acuity and visual field. Quality-of-life measures can be included in selected trials, and the convergent validity of measures should be explored.
2. Quality-of-life measures should not be included in all trials, since their relevance depends on the purpose of the trial.
3. Existing general scales should be considered before developing completely new scales, since new scales have to be validated. Quality-of-life questions can be added to existing scales such as the Arthritis Impact Measurement Scale.

4. An existing general scale, such as the Sickness Impact Profile, should be adapted or extended to enable the quality of life of eye care patients to be compared with that of individuals diagnosed with other diseases.
5. To facilitate reaching agreement about the relative value of investments in eye health care, utility measurements should be made for certain key health states relating to degrees of visual impairment.

Quality-of-life measures can increase the relevance of results of clinical trials to potential patients and the general public. However, they should be incorporated in a way that will not greatly increase the complexity and costs of trials. Further, these measures must enhance rather than detract from the participation and cooperation of both patients and clinical investigators.

# Introduction to the Workshop

Carl Kupfer, M.D.
Director, National Eye Institute

The Workshop on Measuring the Quality of Life of People with Visual Impairment, sponsored by the National Eye Institute on September 13-14, 1988, brought together experts in two fields. Vision researchers met with experts in quality-of-life measurement to exchange information and ideas on how best to assess the impact of eye health interventions on quality of life. Their charge was twofold: 1) to review current issues and knowledge in the field of quality-of-life assessment, with special reference to how they may apply to conditions of the visual system; and 2) to evaluate existing scales for their potential utility in determining the effect on the quality of life of new treatments studied in clinical trials supported by the National Eye Institute (NEI).

The NEI has attempted for many years to measure the cost of visual disorders and to assess the cost-effectiveness of its clinical trials. The Institute has done so by calculating direct medical care costs, such as physicians' fees, hospitalization, and drugs, and by estimating associated indirect costs, such as loss of productivity or the economic impact of disability or premature death. Using this approach several years ago, the NEI estimated that the total economic cost of visual disorders and blindness in the United States in 1981 was about $14.8 billion. The estimate for 1988, after adjusting for inflation, was in excess of $19 billion. Calculating these costs and the savings generated by more effective treatments is one way that policymakers in the Department of Health and Human Services, as well as members of Congress, can ascertain the Nation's return on its investment in eye research.

Savings have been impressive, but they represent only a part of the benefits derived from application of research results to new and better treatments for visual impairment and blindness. Value of treatment should not be measured only in monetary terms. Improvements in a patient's longevity, visual function, or general well-being—his or her quality of life—are among the most important treatment outcomes as well as the greatest potential benefits of any clinical trial.

Yet up to now, vision researchers have found themselves ill equipped to measure improvement in quality of life. The Workshop on Measuring the Quality of Life of People with Visual Impairment was an important step toward supplying the necessary tools. To lead us, we were indeed fortunate to have two outstanding co-chairmen—Dr. Michael F. Drummond, Director of the Health Services Management Centre at the University of Birmingham in the United Kingdom, and Dr. Frederick L. Ferris, III, Chief of the Clinical Trials Branch in the Biometry and Epidemiology Program at the National Eye Institute. Their skillful guidance of the discussions and the excellent papers presented at the workshop, as summarized on the pages that follow, have helped to determine how methodologies for assessing the effect of new treatments on the quality of life might be used in clinical trials supported by the NEI.

# Measuring and Valuing Health States Relating to Visual Impairment: Background and Overview

Martin Buxton

Major developments are taking place in health care evaluation. As interest grows in assessing medical interventions in terms of their impact on quality of life, various clinical research bodies (including many of the major pharmaceutical companies) are convening workshops and seminars on this subject. This interest reflects in part the existence of a now major body of excellent research on health status measurement, including what may be termed "health-related quality of life."

Several recent conferences have reviewed the state of the art. Their proceedings (Portugal Conference, 1987; Palm Springs Conference, 1987) show not only work on methodological development but also increasing use of quality-of-life measures in clinical trials and related studies (e.g., Bombardier et al., 1986; Croog et al., 1986). A separate but associated issue is interest from policymakers and decisionmakers at all levels in obtaining evidence on the relative cost-effectiveness of different medical interventions.

Against this background of increasing interest in evaluating quality of life, those concerned with eye health care are confronted with high incidences of some eye diseases. Estimates of the incidence of visual impairment in communities are notoriously variable. Using figures from a local survey, one study in the United Kingdom estimated that more than 1 of every 100 adults in private households are visually impaired, based on the standards of the World Health Organization (Cullinan, 1977).

With rare but notable exceptions, eye disease is not life-threatening and does not attract attention or resources through mortality statistics. But eye disease does impact significantly on quality of life. Diseases of the visual system are often gradual and degenerative, and frequently they are associated with other diseases in the elderly. Advances in medical science and technology have made it possible to cure, stabilize, or prevent many eye diseases, and many of the programs supporting these advances are not expensive per capita. Health care funding bodies, however, may well consider them expensive in toto.

This paper discusses general considerations concerning the measurement of both visual function and quality of life. It also suggests steps that the National Eye Institute (NEI) could take to develop an instrument to measure the impact of eye health interventions on quality of life.

## Measuring Visual Function

A wealth of good clinical evaluation of eye health interventions now exists, much of it supported by the NEI. The two types of traditional measures used in such studies are clinical manifestations of disease and specific indicators of aspects of visual functioning, such as measures of visual acuity and visual field.

Methods of measuring visual acuity, such as the log of the minimum angle of resolution (MAR), have been used frequently to provide a linear scale (Westheimer, 1979). Other methods have attempted to go one step further and to define levels of visual impairment (based on visual acuity and field) and visual disability (based on ability to perform tasks) (Colenbrander, 1977).

Some studies have attempted to measure the impact of visual impairment on the individual. However, much of the work in this field is entirely clinical, focusing on the eye as an independently functioning organ. Studies often do

not consider the impact of visual impairment on the individual. They may ignore, for example, that the impact is very different if there is sight loss in one eye when the other is unimpaired, compared to when the other eye is blind.

## Measuring Health Status In Relation to Visual Function

A crucial but largely unexplored issue is the relationship between eye function and general health status. No simple, fixed relationship exists between them. The relationship may vary with:

- Function of the other eye
- Age
- Concomitant disease
- Age at loss of visual function
- Psychological reaction
- Nature of social support

If, therefore, the relationship between visual function and general health status is complex, influenced by a range of other factors, studies concerned with visual impairment need to measure directly general health status or health-related quality of life.

## Types of Instruments

The literature on quality-of-life measurement includes a number of problem- or disease-specific instruments, but few have been developed specifically for the field of eye health care. Many do focus on functional disability, including mobility, and, although generally designed to measure physical disability and handicap, some of these instruments may prove sensitive to levels of visual impairment.

Alternatively, general instruments that cover various aspects of quality of life can, at least in principle, be used with any patient group. Subsequent papers in this volume review existing instruments and discuss the relative advantages and disadvantages of disease-specific versus general ones. There are three main groups of general instruments:

- *Health profiles* incorporate a number of different dimensions of health, e.g., energy, social interaction, mobility, or emotional reactions, and produce a measure for each dimension. Profiles provide an overall picture of health status.
- *Health indices* attempt to bring these various dimensions together and express health status as a single point on an index.
- *Utility measures.* Some health indices adopt a utility concept, which attempts to show the relative value of different points on a single index, using an interval or ratio scale.

These approaches can be characterized as reflecting three different processes. The first is one of description or classification: How do we distinguish significantly different health states? The second is one of ordering or ranking these states, or measuring them in terms of natural units: Which states do we prefer to others? The third approach is a process of valuation or ascribing utilities: How much do we value one health state relative to another?

## Characteristics of Instruments

If measures are to be used, what attributes do they need to exhibit? Feinstein (1987) has coined the term "clinimetrics" to describe the characteristics of clinical measures. His nomenclature of attributes offers a useful framework that escapes from the traditional approach based on characteristics of measures that can be checked against a laboratory "gold standard." He suggests instead that measures need to satisfy three main groups of criteria:

- Sensibility, including appropriateness of purpose, comprehensibility, replicability, suitability of scaling, face validity, content validity, and ease of use.
- Consistency, which considers test-retest reliability and inter- and intra-observer variability.
- Validity, which includes quantitative checks on construct and criterion validity.

Some of the best available measures do come close to passing these tests. Indeed, a number of the new health status measures may meet these criteria better than some traditional clinical measures.

## Contexts of Use

Well-validated health status measures could be valuable in a range of situations, such as clinical trials, trials specifically designed to measure quality of life, cost-effectiveness studies, multipurpose experimental studies, and service evaluation and monitoring. Each of these different contexts imposes somewhat different requirements and constraints on the measures used for quality-of-life assessment and the technical processes of administering the instruments or eliciting the necessary information. The choice of measures, therefore, may be different in each case. Nevertheless, there is an important role for quality-of-life measurement in each of these contexts.

## Purposes of Measures

In all applications, the purpose of obtaining quality-of-life information should be considered. Quality-of-life measures can be used in two ways:

- To measure change or difference within a particular trial, study, or monitoring exercise. For this purpose, a discriminatory measure is needed to distinguish among different groups or different points of time within studies. Discriminatory power is the key factor.
- To relate to data obtained outside the trial. In such cases, the quality-of-life measures are required for external comparisons to relate the quality of life of patients from a particular study to the broader spectrum of existing knowledge of the quality of life experienced by quite different groups. Here, common measurement and comparability are the main requirements.

## Conclusion

An in-depth review of this topic is presented in Appendix A (Ferguson et al., 1988). The review suggests four tentative conclusions about ways that the NEI should direct its interest in quality-of-life measurement.

- Some existing general instruments should be used in selected clinical trials, closely monitored by those familiar with use of the instrument in other patient or population groups. The Sickness Impact Profile (Bergner, 1981) is an obvious candidate.
- The NEI should set up an expert group to consider what factors and dimensions need to be incorporated into a specific instrument for the eye health care field—perhaps building on existing functional assessment scales. Small-scale experiments might be conducted to begin to indicate the potential value of a specific instrument.
- In the light of the results of these first two steps, the NEI should consider whether one or both approaches should be pursued and what developmental work needs to be performed.
- As a separate exercise, utility values for a number of key states of visual impairment should be established. These states could be chosen on the basis of existing quality-of-life evidence and professional judgment. Empirically determined utility values for such states would then provide a starting point for establishing comparisons of the cost-effectiveness of eye health care interventions relative to other uses of health resources.

## References

Bergner M. The sickness impact profile: Development and final revision of a health status measure. *Medical Care* 19:787-805, 1981.

Bombardier C et al. Auranofin therapy and quality of life in patients with rheumatoid arthritis. *The American Journal of Medicine* 81:565-577, 1986.

Colenbrander A. Dimensions of visual performance. *Transactions of the American Academy of Ophthalmology and Otolaryngology* 83:347-350, 1977.

Croog SH et al. The effects of antihypertensive therapy on the quality of life. *New England Journal of Medicine* 314(26):1657-1664, 1986.

Cullinan TR. The Epidemiology of Visual Disability: Studies of Visually Disabled People in the Community, Research Report No. 28, Health Services Research Unit, University of Kent, Canterbury, 1977.

Feinstein AR. Clinimetrics, New Haven: Yale University Press, 1987.

Ferguson BA, Buxton M, and Drummond MF. Measuring and Valuing Health States Relating to Visual Impairment: A Review of Literature, Concepts and Methods, HERG Research Report No. 3, Brunel University, Uxbridge, 1988.

McDowell I and Newell C. Measuring Health: A Guide to the Rating Scales and Questionnaires, New York, Oxford University Press, 1987.

The Palm Springs Conference. Advances in health assessment. *Journal of Chronic Diseases* 40(suppl 1), 1987.

The Portugal Conference. Measuring quality of life and functional status in clinical and epidemiological research. *Journal of Chronic Diseases* 40(6), 1987.

Westheimer G. Scaling of visual acuity measurements. *Archives of Ophthalmology* 97:327-330, 1979.

# How Visual Impairment Affects Quality of Life: Experiences with Quality-of-Life Measurements in Clinical Ophthalmology

Peter Bernth-Petersen, M.D., Ph.D.

In many patients with visual impairment, it is difficult to equate visual acuity measurements with their level of visual functioning. Visual functioning can be defined as the performance of everyday tasks that are dependent on vision, such as coping with one's job, housework, self-care, and mobility. To evaluate the benefits of treatment, standardized measurements of levels of visual functioning are needed to supplement measures of visual acuity. This paper describes one such measurement, the Visual Functioning Index, which was developed in Denmark and had an important impact on that country's policy concerning cataract surgery.

## Visual Functioning Index

The Visual Functioning Index measures visual functioning through self-assessment. It provides information on direct visual limitations, mobility limitations due to visual impairment, and social role limitations, including need for help due to visual impairment. The construct validity of the Index has been found acceptable. Classification according to visual function is possible by applying a specific scoring system to the Index.

Easy to use in patient interviews, the Visual Functioning Index was applied to a group of cataract patients in Denmark. Results suggested a positive correlation between visual acuity and visual functioning, but the relationship was not always consistent. The Index showed that about 30 percent of bilateral cataract patients had very poor visual functioning at the time of hospitalization, imposing severe limitations on their social independence. This finding was used successfully as an argument for cataract surgery at an earlier stage of the disease than was usual in Denmark.

All patients underwent cataract extraction. The outcome of cataract surgery was assessed 1 year later by measuring visual acuity and by applying the Visual Functioning Index, the scoring system, and other quality-of-life measures. Postoperative findings were compared to preoperative findings. The results, which were favorable, contributed to allocation of more resources to cataract surgery in Denmark in the early 1980s.

## Conclusion

The use of the Visual Functioning Index in this study allowed a profound quantitative and qualitative evaluation of the benefit of cataract treatment. This evaluation included rehabilitation issues as well as occupational, social, and economic aspects of quality of life.

Using quality-of-life measurement (QLM) in clinical ophthalmology had these benefits:

- QLM provided detailed information about consequences of eye disease.
- QLM considered broader aspects of the benefits of treatment of eye disease.
- QLM provided data that were understandable to non-ophthalmologists (patients, politicians, health care decisionmakers), thereby facilitating communication and helping to influence decision-making.
- QLM provided the basis for rational eye care planning.

Experience with the Index has also led to these conclusions:

- QLM should be based on self-assessment. The scales should not be too complex; patients should find them easy to use and understand.

- Different tasks need different instruments. Constructing a single QLM scale for use in clinical trials of different eye diseases would be extremely difficult.

## Bibliography

Bernth-Petersen P. Visual functioning in cataract patients - Method of measuring and results. *Acta Ophthalmologica* 59:198-205, 1981.

Bernth-Petersen P. Outcome of cataract surgery I: A prospective observational study. *Acta Ophthalmologica* 60:235-241, 1982.

Bernth-Petersen P. Outcome of cataract surgery II: Visual functioning in aphakic patients. *Acta Ophthalmologica* 60:243-251, 1982.

Bernth-Petersen P. Outcome of cataract surgery III: Influence of age, macular disease and type of correction. *Acta Ophthalmologica* 60:455-460, 1982.

Bernth-Petersen P. Outcome of cataract surgery IV: Socio-economic aspects. *Acta Ophthalmologica* 60:461-468, 1983.

Bernth-Petersen P and Sorensen T. Intraocular lenses versus extended-wear contact lenses in aphakic rehabilitation. *Acta Ophthalmologica* 61:382-391, 1983.

Bernth-Petersen P and Ehlers N. Cataract extraction in the Nineties. *Acta Ophthalmologica* 61:392-396, 1983.

Bernth-Petersen P. Cataract surgery - Outcome assessments and epidemiologic aspects. *Acta Ophthalmologica* 63 (suppl 174):1-47, 1985. (A summary of the above mentioned papers.)

Scherfig E and Bernth-Petersen P. Vitrectomy in diabetics. Outcome assessments 2-5 years postoperatively (in press).

# Discussion

**DR. ATWELL:** You said in your conclusion that a general measure would be difficult to develop, and yet the kinds of questions that you used sounded fairly general and applied to most kinds of activities. What measures would be inapplicable to other eye diseases, such as macular degeneration?

**DR. BERNTH-PETERSEN:** To get the full picture, we had to supplement with other questions outside the Index. A scale that covered all aspects of vision would be extremely complicated. For example, if you wanted specific information about treatment for macular degeneration, you would ask if the distortion of vision had disappeared. You wouldn't ask a cataract patient that because he wouldn't have the symptom. You might use a basic scale, but you could not get the specific information you needed in a clinical study if you didn't ask for the specific symptoms of that disease.

**DR. ATWELL:** My question is about the relationship between the symptoms of disease and the patient's general functioning, which seemed to be the emphasis in your study. For example, you asked about mobility and reading. To what extent are the very specific symptoms that may accompany specific diseases qualitatively important for the general or everyday functioning of the individual?

**DR. BERNTH-PETERSEN:** I don't know, but many people with senile macular degeneration have normal visual functioning, so you won't get any exact information by using a general scale there. In the beginning at least, they have normal visual acuity and they function normally. I think you would need a disease-specific instrument to assess the patients' visual impairment and to measure the treatment effect, if any.

**DR. PATRICK:** With macular degeneration, the prognostic implication of the disease has importance that would not be picked up by a function status measure like his Index.

What more did you get out of the Index, other than finding that 74 percent were satisfied with the outcome of their cataract surgery?

**DR. BERNTH-PETERSEN:** We were able to show the severity of the problem of cataracts and to describe the functioning preoperatively and postoperatively.

**DR. PATRICK:** The visual acuity correlated very well with specific items on your index, but once you made an overall score, you found people with high visual impairment and good function, and people with low visual impairment and poor function. Your tables on

your individual items were all quite consistent, but once you aggregated your score, you had some problems. Would you agree?

*Our results confirmed indirectly the need for measurements other than those for visual acuity to get an exact description of what people can or cannot do.*

**DR. BERNTH-PETERSEN:** No, I wouldn't. I think that if you had a perfect correlation between visual acuity and functioning, as measured by scoring, there wouldn't be any reason for measuring visual functioning in another way than visual acuity.

We do not get enough information from visual acuity measurements. Our results confirmed indirectly the need for measurements other than those for visual acuity to get an exact description of what people can or cannot do.

**DR. FERRIS:** Presumably everybody was able to read. I don't know about driving a car and so on. Do you also ask whether they previously functioned at that level?

**DR. BERNTH-PETERSEN:** When we asked the questions, we specifically stated that the limitations addressed should be due to vision. So, I think we did eliminate the problem you mention.

**DR. FERRIS:** So everybody that answered, "No, I don't drive a car," previously did drive a car?

**DR. BERNTH-PETERSEN:** Yes. We asked patients whether their self-transportation capabilities were limited because of their vision.

**DR. MURPHY:** How do you adjust for individual expectations and needs? Obviously, the expectations and needs of an engineer who's 32 years old might be quite different from those of an 82-year-old woman in a nursing home. How do you adjust for that?

**DR. BERNTH-PETERSEN:** I don't.

**DR. BERGNER:** What was the age of the patients?

**DR. BERNTH-PETERSEN:** The average age was 72.

**DR. KLEIN:** Opthalmologists are performing more cataract surgery on patients with better visual acuity. How do you control for what patients have been told prior to their surgery and how this biases their answers?

**DR. BERNTH-PETERSEN:** Expectations were very great in the early 1980s. But of course, when designing a scale, you should take that into consideration, too. I interviewed and informed all the patients myself, and I took care not to influence their answers. In this way, I tried to minimize these kinds of bias.

**DR. BAROFSKY:** Although you said it was an ancillary issue, I think it is in fact a primary issue that you accomplished your purpose with this assessment, which was to change policy relative to cataract surgery in Denmark.

**DR. BERNTH-PETERSEN:** My main interest was in learning whether patients had any benefit at all from this treatment. We heard a lot of complaints in those days. The patients couldn't get used to the aphakic glasses. The patients wouldn't wear them, and they used the other eye with the cataract, and so on.

One motive for going into this study was to determine if aphakic function could be improved. In a later study we showed that we could not only improve the function, but we could also improve the patient's contentment by using intraocular lenses. And this was part of the purpose. The results of this study were rather surprising at that time. Many of the ophthalmologic departments were like sleeping beauties in the early 1980s, and I think this gave them a sort of kick.

**DR. LOHR:** Did you consider having any items other than reading for measuring near-vision functioning? Did you consider other activities such as sewing, handwork, or cooking that might require a higher level of functioning than reading does?

**DR. BERNTH-PETERSEN:** We asked about all activities, but not specifically to measure near-vision functioning other than reading.

**DR. LOHR:** What prompted me to ask specifically about sewing and handwork as leisure activities distinct from reading is that their loss can be a serious blow to some patients. Questions about these activities could give a different dimension to the restoration of functioning. Did you consider testing some other items to see if they would give you a broader, different measure?

**DR. BERNTH-PETERSEN:** Yes, but we didn't do so.

**DR. FERRIS:** Probably a number of the ophthalmologists here are trying to learn about how one uses these scales. Most of my blind patients tell me that even though sightless, they can cook. Sure they can cook, but it would be a lot easier for them to cook if they could see. How do you build these activities into the scale?

**DR. BERNTH-PETERSEN:** The results for indoor orientation were quite independent of visual acuity. The self-care question about bathing, dressing, eating, was not very specific either, because these functions are rather independent of visual acuity. These items, therefore, were only given little weight in the system of appropriate scoring as applied to the Visual Functioning Index.

# How Visual Impairment Affects Quality of Life

Gordon E. Legge, Ph.D.

This paper comments on quality of life with visual impairment from two perspectives: that of a person with impaired vision and that of a researcher who studies the effects of visual impairment. It discusses the advantages of objective over subjective measures and goes on to discuss specific ways in which visual impairment affects quality of life, with special consideration of those elements that may be measured objectively.

## Objective Versus Subjective Measures

In their paper, "Measuring and Valuing Health States Relating to Visual Impairment: A Review of Literature, Concepts, and Methods," Appendix A of this document, Ferguson, Buxton, and Drummond describe the limitations of acuity as an outcome measure and recommend development of quality-of-life measurement (QLM). In considering QLM, however, it is important to understand the limitations of subjective measures.

I would like to give you an example from my own research. We conducted a study measuring the effects of the color of text on reading performance; that is, we measured reading speed as a function of either the color of letters on a black background, or the color of the background. In all cases we equated the luminance of the wavelengths of the colors used, so that wavelength was the only independent variable.

Subjectively, almost all of our subjects, normal and low vision alike, had very decided preferences for the colors that they felt that they would read best with. Objectively speaking, their reading speeds in most cases were unaffected by the wavelength. There were a few deviations from that generality, but for the most part, wavelength had no effect. In this case, subjective preference did not correlate closely with the objective performance measurements.

QLM is subjective and relies on self-evaluation or possibly the appraisal of outsiders. It is one step removed from objective performance measures such as acuity, contrast sensitivity, or reading speed. QLM thus may misrepresent true visual abilities.

It seems likely that subjective QLM will be more sensitive to two temporal variables—age and period of impairment—than objective performance measures. The quality-of-life estimates of a student and a retired person with the same eye condition might be very different. Quality-of-life estimates also will depend on how long and how well a person has adapted to a condition of visual impairment. In evaluating a clinical trial, who makes the quality-of-life judgment and at what stage after the treatment?

Performance measures and subjective QLM may serve different functions. Objective performance measures probably are linked more closely to a person's biomedical condition. If properly designed, they are insensitive to nonphysiological factors. Subjective measures can be more sensitive to a person's perceptions of his or her health state. They depend not only on one's current biomedical condition but also on one's prior condition and expectations, and on personality and social and economic variables.

A principal research goal of clinical trials is to determine the effects upon vision of some procedure. The most direct evaluation should be based on objective performance measurements. Subjective evaluation is important in determining acceptance of the procedure and its practical impact, but it brings into play variables that may lie outside the scientific scope of the clinical trial.

## Ways in Which Visual Impairment Affects Quality of Life

When you ask someone what bothers them about being visually impaired, they are likely to mention one or more of the following difficulties with everyday life:

- Reading — difficulties that make it impossible, slow, or tiring.
- Driving — loss of license; difficulties caused by glare, darkness or problems with peripheral vision; inability to read street signs.
- Getting around on foot — fear of bumping into things, tripping, falling on stairs; difficulty at night or in bright sun.
- Watching TV — necessity of sitting close to the screen or difficulty in following the action.
- Pursuing hobbies and sports — inability to play golf or tennis, do needlepoint, or pursue other activities.
- Suffering social embarrassment — difficulty in recognizing and acknowledging friends; awkward encounters with strangers; difficulties in noisy group situations such as parties.
- Enduring lack of privacy and dependency — necessity to seek help with correspondence, sorting socks, or cooking.
- Having concerns about appearance — self-consciousness about using glasses, magnifiers, white canes, guide dogs, or other aids.
- Bearing a stigma — aversion to being classified as legally blind.

Some of these factors are amenable to objective performance measures. For example, in my laboratory we currently are addressing the question of how various states of low vision affect reading performance. Similar work can be or is being done with mobility, driving, and face recognition. Concerns about problems such as stigma and social embarrassment may require a more subjective analysis.

From a broader perspective, visual impairment also can affect quality of life in long-term ways, including these:

- Education — closes some options and slows progress.
- Career — closes some options; reduces prospects for promotion and reduces job mobility; job placement may require special adaptations or accommodations by employers which tend to lock the employee into a position.
- Mobility — reduces access to driving, public transportation, and pedestrian travel.
- Social life — reduces social integration with mainstream society.
- Recreation — reduces range of alternatives for leisure and recreation.
- Independence — promotes greater dependence on family, friends, or social services personnel.
- Flexibility — makes everyday problem solving more difficult and time-consuming because of the reduced number of options available.

Some of these broader problems also might be amenable to objective analysis. For example, we might ask how employment, job income, or educational advancement relate to visual acuity or visual field. Does acuity or some other biomedical measure predict whether elderly people live independently or in nursing homes?

## Conclusion

Using subjective measures of quality of life in evaluating the outcome of clinical trials is intriguing and provocative, but this approach should supplement, not replace, improved objective measures of performance. Reservations concerning quality-of-life estimates relate to their stability and their role in evaluating the visual consequences of clinical procedures, as opposed to economic or social consequences.

# Discussion

**DR. BERNTH-PETERSEN:** If we consider objectivity and subjectivity in clinical ophthalmology, visual acuity is partially subjective; visual field testing is partially subjective; and contrast abilities are also partly subjective. In fact, there is a very large amount of inter- and intra-observer variation in those tests. Two British investigators, Gibson and Sanderson, reported that their most consistent findings when dealing with partially sighted patients were the complaints—not visual acuity and not visual field testing.

**DR. LEGGE:** I don't know the study you referred to. I do know, however, that properly designed acuity or contrast-sensitivity measures have rather small variability. Therefore, while I agree that there is a subjective component in any sort of behavioral measure like that, a well-designed test can provide, at least in my view, a good measure. I have less experience with visual field testing, and I gather this is more an art form than an objective measure. But someone else probably knows much more about that.

I think that there's a signal-to-noise ratio problem here. Any measure of anything will have variability in it. The question is, what is the signal-to-noise ratio? I think that in a properly devised acuity or contrast-sensitivity test, the signal-to-noise ratio can be high. It's arguable what you do with those results, but I think those kinds of measures can be made fairly cleanly.

**DR. WARE:** At first I thought I disagreed with you because you seemed to be coming down pretty hard on subjective measurement. But you added that clearly one shouldn't replace the other.

Then I began to agree with you a great deal. The fact is, you can't measure quality of life without subjective measurement. You gave the example of one person enjoying a particular color mix more than another one. The issue is, what is the utility of that enjoyment?

If the argument against subjective measurement is that it's less reproducible or that it doesn't meet the standards of measurement as well, that might have been true 50 years ago, but it isn't necessarily true today, with good application of techniques. Are you really criticizing the notion of very subjective concepts, like satisfaction and so on?

**DR. LEGGE:** I'm doing a number of things. A little bit of it is devil's advocacy. But my bias is toward performance measures. My prejudice is that performance measures will give us a more accurate index of the visual consequences of a procedure. That's the main purpose of some clinical trials—to improve vision. I would argue that if we want to evaluate how much vision is improved, we should test vision. We should test it in as objective a manner as possible.

I agree that subjective measures capture things that are not captured by performance measures, but I would contend that the subjective measures take into account not only visual factors, but many other factors, whereas performance measures can be confined more to measuring changes in vision.

**DR. ATWELL:** I think that last point is one we should probably bear in mind. We're talking about two categories of measures: One is the performance measure of vision, and the other is what difference it makes in functioning.

These are complementary measures; we're not talking about one necessarily being more valuable than the other. It depends upon utility. In one case, you're going to measure the performance that people can achieve, and in the other, how happy or satisfied they are with what they can achieve, how important it is to them. Some people may not care whether they can read, but they may care a lot whether they can look at television across the room.

*The fact is, you can't measure quality of life without subjective measurement. You gave the example of one person enjoying a particular color mix more than another one. The issue is, what is the utility of that enjoyment?*

**DR. LEGGE:** In reading the background paper prepared by Ferguson, Buxton, and Drummond (Appendix A of this document), which was very informative to me, I was concerned that the recommendations and the content of the paper did not address alternative, additional, or improved performance measures. I agree that there are limitations in existing measures and performance measures. I think there's room for improvement there, as well as in the development of subjective measures.

**DR. BERGNER:** I think we sometimes talk about performance in two different ways: "Can you do something?" and "Do you do something?" I could imagine a clinical trial in which you improve visual acuity, but in terms of what happens to the patients—how they carry on their lives, and whether they make any changes in their performance—there may be no improvement. So you may put patients through a procedure that may not affect how they carry on their lives at all. I assume that this is one of the questions one would like to address.

**DR. ATWELL:** There's actually a classic case of just this distinction that's being researched by a former colleague of Dr. Legge, Denis Pelli, who did research on the visual requirements of mobility. The actual visual requirements for being able to do something like go to the supermarket and find a particular item on the shelf were very low; people could do it, but in fact they didn't. Pelli ultimately concluded that in order to understand why, you had to get into other measures, such as motivation, acceptance, and peer expectations. That's the direction research has to take. It really was not a visual problem per se, even though that would have been the original assumption. I think that's an important distinction.

**DR. LEGGE:** One question that occurred to me is, "What is the mandate of those conducting the clinical trial?" If they restore perfect vision to someone, is that their objective, or is their objective to get the person out of the house? Maybe they're housebound for some other psychological reason? Maybe they'll never read or drive no matter how perfect their vision.

It seems to me that this is not the immediate concern of those conducting clinical trials of a particular ophthalmological technique.

**DR. McDOWELL:** But in a way, that's rather a sad attitude to take. It could be misinterpreted as the perennial tendency to divide people up into little pieces; eyes or feet or whatever. No one quite gets around to fitting all the pieces together again, other than perhaps some nice benign rehabilitation people.

I would hope that one of the messages from this conference is that we should look not only at restoring patients' visual acuity but also at restoring some of the things that they would like to do.

**DR. FERRIS:** In the Early Treatment Diabetic Retinopathy Study, we incorporated a subjective component for some of the reasons that have been mentioned. Also, some of us who had done photocoagulation had seen patients who, although their visual acuity improved, preferred their other eye, which had lower visual acuity but hadn't had laser treatment. It was interesting that the results of the visual acuity test were quite close to the results of the subjective examination that asked questions like: "Do you prefer your right eye to your left eye?" "If you had to drive a car with just your right eye, would you?"

That's one of the few things for which we've been criticized in that study. Our critics said we should never have asked such questions. So, we're living in a world of severe skepticism. People are concerned about problems of bias in subjectivity and about patients' wanting treatments to work.

*I would hope that one of the messages from this conference is that we should look not only at restoring patients' visual acuity but also at restoring some of the things that they would like to do.*

Of course we were cognizant of these issues at the start. On the other hand, we thought we'd take a first shot at trying to ask subjective questions.

I'm a little surprised to hear you say that you think that visual acuity results are not very reproducible. At least in the clinical trials where we've done reproducibility studies, we wind up with 90-plus percent within one line when we replicate visual acuities of the same patients on the same day. From that perspective, we concluded that the test is fairly reproducible. I think all of us recognize the importance of visual acuity, but it doesn't say everything there is to say about visual functioning.

In the questionnaire-based studies that I've been involved with, I have worried a great deal about the answers obtained. For example, in asking people, "Can you recognize someone across the street?" there are examples where someone has 20/20 acuity, so there isn't any good reason why they shouldn't recognize someone across the street, but they answer the question, "No, I can't recognize people across the street." There may be many reasons why they answer the question that way.

The problem has always been in interpreting these numbers. For example, if you said that people who can't recognize someone across the street must have impaired visual acuity, and you then try to extrapolate that number, using the National Health Interview Survey, to reflect the population in the United States,

it would turn out that there are as many people having 20/20 acuity saying they have impaired vision as there are people that really have impaired vision.

So, whatever measures we develop must be not only reproducible but also valid.

**MR. McMANUS:** I think that earlier it was stated that quality-of-life indicators can be used in two different ways—narrowly as measures of effectiveness in clinical trials or more broadly for cross-program, cross-purpose evaluation.

How do you measure the worth of putting some economic investment in an eye clinical trial versus a cohort study for heart disease or cancer? That's the real practical problem that we at NEI are now facing. How do we compete for resources with these other large-scale programs at the NIH? Is there any objective way that we can compare the outputs of our studies with the outputs of other studies? We and our ophthalmology colleagues feel that the quality of life of eye patients is much more improved than those in any other area. I mean, just on the face of it, you should put your investment here. It would be nice if we could find a way to quantify that, so that we can compare the value of different programs of different Institutes.

**DR. DRUMMOND:** In summing up, it would probably be easier to identify the issues we didn't cover than those we did in our paper, "Measuring and Valuing Health States Relating to Visual Impairment: A Review of Literature, Concepts, and Methods." The first thing I should do is correct one possible misinterpretation of the paper. That is, I don't think we were saying that there should be wholesale substitution of the kind of measures now used with the more subjective quality-of-life measures.

However, in the real world, when you are conducting trials in particular, there are limitations in terms of budget and the time that you can get respondents or patients to participate. So there is some question of priorities in selecting methods of assessment.

In addition, we need to examine further the extent to which general quality-of-life measures can encapsulate some of the things Gordon Legge was talking about. I must admit, when he was listing some of the aspects of visual impairment, I was mentally mapping those onto the kinds of scales that I know about. There seemed to be quite a lot of overlap.

On the other hand, the discussion touched on different kinds of measures. Some assess the consequences of visual impairment, which in some sense could be measured objectively, while others measure the impact of visual impairment on quality of life, which may be quite a different thing.

Another issue that I think we'll have to come back to is the purpose of clinical trials. Is it to show improvements in visual functioning? Or is it to show the benefits of treatments in the eye health care field in a way that would enable them to be compared with interventions in other fields of health care?

# Measures of Outcome in Current Clinical Trials of Eye Care

Ronald Klein, M.D., M.P.H.

Many outcome measures used in clinical trials of eye care measure the visual consequences of treatment. For example, a review of 14 current clinical trials sponsored by the NEI found that all measure visual acuity as an outcome, six look for changes in visual field, three evaluate color vision, two assess contrast sensitivity, one includes the glare test, and one tests for binocular fusion (Table 1).

This paper discusses outcome measures of visual acuity, contrast sensitivity, visual fields, and color vision; it does not discuss other outcomes such as changes in the intraocular pressure or in the severity of diabetic retinopathy. Each description covers basic principles underlying the measures, test methods, and the relevance of the measurement to day-to-day functioning.

**Table 1**
**Visual Function Tests Used in 14 NEI-Sponsored Clinical Trials**

| Tests | Number | Percent |
|---|---|---|
| Visual acuity | 14 | 100 |
| Visual field | 6 | 43 |
| Color vision | 3 | 21 |
| Contrast sensitivity | 2 | 14 |
| Glare testing | 1 | 7 |
| Binocular testing | 1 | 7 |

## Visual Acuity

### *Principles*

Visual acuity may be the most important indication of day-to-day function. It is frequently used as an outcome measure because it is relatively easy to assess visual acuity with the standardized objective tests available, the tests usually cause no discomfort to the subject, and the results can be subjected to statistical analyses.

Visual acuity is measured as the visual angle at which components of an object can be appreciated. In tests of visual acuity, letters are designed so that at a specified distance, the whole letter subtends 5 minutes of arc at the eye and each detailed part of the letter subtends 1 minute of arc. Thus, for Snellen letters, 20/20 (6/6) means that at a distance of 20 feet, the subject can read letters composed of high contrast black lines on a white background, with the width of each line, white and black alike, subtending one minute of angle at the eye; 20/40 requires a letter whose components each subtend 2 minutes of angle at a distance of 20 feet from the letter; 20/10 requires a letter whose components each subtend 1/2 minute of angle at a distance of 20 feet from the letter.

Looking at this phenomenon another way, a letter must get larger, as its distance from the eye increases, to continue to subtend an angle of 5 minutes at the eye. Thus, an eye with 20/20 visual acuity can see the large "E" at 200 feet, while an eye with 20/200 visual acuity can see it only at 20 feet. Snellen fractions, such as 20/40, can also be expressed as a decimal fraction (0.5), a minimum visual angle resolved (2.0), and a log scale (0.3).

### *Tests*

Standard test conditions are set before the beginning of each trial. For example, in visual acuity testing one must specify the illumination necessary to attain a specific contrast (usually about 85 percent or greater between the letter and its

background); one must also specify the distance between the subject and the chart (usually between 10 to 20 feet).

Test charts are specially designed for many of the studies. The LogMAR Chart has been used in the Early Treatment Diabetic Retinopathy Study (ETDRS) and other trials and has as relative advantages the same number of letters on each line, regular progression of letter size from line to line, lines using a combination of letters so that they are of approximately equal difficulty, a difference between lines of 0.1 LogMAR, and letters that have 0.02 LogMAR units.

Standardized refraction schemes have been developed to measure visual acuity as corrected to the best possible level. Visual acuity requires not only resolution but the ability of the subject to detect the letter, recognize it, and vocalize his or her response. Testing conditions are very important for the measurement of visual acuity.

### *Functional Importance*

Various terms used to describe decreased visual acuity—such as legal blindness, minimal visual impairment, no light perception—have importance for job performance, activities of daily living, and other aspects of day-to-day functioning.

## Contrast Sensitivity

### *Principles*

Perception of images in the environment requires the visual system to process differing contrasts across a wide spectrum of spatial frequencies. There is a point at which contrast—the difference between the energy reflected by the object and that reflected by the background—cannot be detected. This is called threshold contrast. The eye is not sensitive enough to appreciate different letters and background below threshold contrast. Contrast sensitivity is defined as 1/threshold.

Contrast threshold is a function of three factors: wavelength of light, background brightness, and the part of the retina examined. Contrast sensitivity may add further information about the ability to perform visual tasks under varying conditions. The effect of both contrast and visual angle (spatial frequency) on visual resolution can be quantified using contrast sensitivity gratings, such as sine wave gratings.

### *Tests*

One such system, the Vision Contrast Test System, employs a series of sine wave gratings of different spatial frequencies, contrasts, and tilts. Contrast decreases in 0.2 log unit steps from a high of 33 percent to 0 percent; the spatial frequency varies from 1.5 cycles/degree to 18 cycles/degree. The test is usually performed at a distance of 3 meters. This and other tests of contrast sensitivity are not yet as standardized as tests of visual acuity.

### *Functional Importance*

Despite restoration of 20/20 acuity, patients with optic neuritis may complain of visual disability. This has been demonstrated in studies in which patients have complained of problems with color perception, brightness of objects, depth perception, vision needed for doing household chores, and other problems with sight. In such patients, abnormalities have been found in tests of contrast sensitivity as well as in several others, including tests of color vision, perimetry, stereoacuity, and light brightness.

The subjective functional visual impairment of patients with optic neuritis correlated better with contrast sensitivity than with either perimetry or color vision. The functional importance of contrast sensitivity in other diseases that affect visual function, such as open-angle glaucoma and age-related macular degeneration, should be studied further.

## Visual Field

### *Principles*

Every point within the visual field has a visual threshold, which is the weakest test stimulus that is viable in a particular location under the specific test conditions. A frequently referred to concept is the "hill of vision."

### *Tests*

The visual target is usually a white spot of given size and intensity; movement and background illumination are specified. The threshold for target detection depends on the luminance of the

background, luminance of the target, and size of the target. Those administering the test look for the weakest white spot that can be seen in each region of the field, usually adjusting size and intensity.

There are two ways to measure visual field: static threshold perimetry, in which the location is fixed and the white spot is presented at various intensities; and kinetic threshold perimetry, in which the stimulus is moved from an area where it is not seen toward an area where it is seen.

### *Functional Importance*

Loss of central field, as with macular degeneration, is associated with loss of reading vision and ability to see faces. It may be corrected with visual aids. Loss of peripheral field, as with glaucoma or retinitis pigmentosa, also is disabling and may be more difficult to correct with visual aids.

## Color Vision

### *Principles*

Color blindness, or more precisely, color deficiency, comes in three forms: protanopia, insensitivity to red, in which the subject confuses red and green; deuteranopia, insensitivity to green, which has the same result; and tritanopia, insensitivity to blue, in which the subject confuses blue and yellow. These are relatively rare. Anomalous color vision results from a deficiency in cone pigments leading to difficulty telling saturated colors apart; the subject confuses pastel colors. Acquired color vision abnormalities, such as problems with blue and yellow, occur in diabetic retinopathy.

### *Tests*

A number of tests, including the Farnsworth-Munsell 100 Hue test and Ishihara plates, measure color vision.

### *Functional Importance*

Most persons with changes in color vision do not have noticeable functional changes. Artists and others who work with color are the exception.

# Discussion

**DR. BERGNER:** I have a comment about the functional implications or impact of color blindness. It's very difficult, if not impossible, to pass a medical school histology exam if you're color blind, and you have to learn from the shapes of the cells rather than the way in which they're stained.

**DR. KLEIN:** Again, it's the situation that counts, and that's a good example. But often, a lot of people have learned to know where the red is and where the green is and they get by very well. In general, except for artists and people with specific needs, color vision has less weight. If we want to weight some of the functional parameters, visual acuity and visual field are the most important in my opinion.

**DR. PATRICK:** I was very interested in the LogMAR charts. Are they being used only in clinical trials, or are they going to replace Snellen charts?

**DR. KLEIN:** I can't answer that question. They're being used in many clinical trials and in many epidemiologic studies. I don't know how quickly they're going to replace the Snellen chart.

It depends again on the needs. In an ophthalmologist's office, the Snellen, if used under the right conditions, is a perfectly legitimate chart. But if it's for trial purposes, the LogMAR chart offers a number of advantages. Rick Ferris designed the LogMAR chart for use in the ETDRS and has written extensively on it. He might want to comment.

**DR. FERRIS:** There surely are some ophthalmologists outside of clinical trials that are using the LogMAR chart because they want a better sense within their own office of how individuals are changing over time. I think the Snellen style chart has lasted for years because it can provide a perfectly good representation of about where the patient's vision is. It gives you the ball-park measurement, and that's what most ophthalmologists are interested in, rather than in small changes. Small changes over time may be true changes, or they may be changes in the way the test was done or reflect other differences. Without strict attention to those details, a change of a line or so wouldn't be considered as important.

It was clear when we got involved in doing these studies that we were interested in these small changes, in what our level of resolution was with the test, and in looking at changes in visual acuity over time. It was extremely difficult to get this information using the Snellen chart. There are a varying number of letters on different lines, there are different degrees of separation between lines.

*When doing our large population-based studies we wanted to add quality-of-life measures. But we found that some are quite time-consuming, and that brought up the same problems: what to put in, who to test it on, and whether the costs are justified by the benefits.*

**DR. PATRICK:** There's a wonderful analogy in this for quality-of-life measures. A Snellen quality-of-life measure may be adequate in clinical practice, whereas in clinical trials, it may be better to use a LogMAR-type measure that's terribly sensitive to minor changes related to visual acuity or visual field or color.

The measurements, whether objective or subjective, bring up the same issue. We need to decide whether we need a LogMar or a Snellen.

**MR. BUXTON:** Do you think we have any LogMAR quality-of-life instruments? That's a serious question.

**DR. PATRICK:** Yes, using the same rigor of test administration for visual acuity, standardization of item content, and administration under perfect lab conditions, we will get similar levels of reproducibility in some quality-of-life measures. But we seldom treat quality of life the same way we treat visual acuity. We don't measure it with the same care.

**DR. BERGNER:** But I heard you say that in some cases, we don't measure visual acuity with all that much care, either. I have two serious questions; first, you said one of the reasons that visual acuity is measured is that it's easy to measure. I don't think you defined what you meant by "easy to measure."

To me this says that the test itself is cheap; that is, all I need is a Snellen chart and I just have to tack it up on a wall. It doesn't require very much for me to do it. I need to have some light and a room. The time that it takes to do it is very limited—5, 10, 15 minutes—and I don't need a trained operator to do it. I can train somebody easily to do it. So that it has all of the aspects that make people very happy in clinical trials. You don't need special equipment or the special equipment is very inexpensive. The person who administers it can be very inexpensive, and the time it takes is not very great. So the "easy to" measure really turns out to be a dollar issue, it appears to me.

**DR. MURPHY:** Cost is an issue, but time is also an issue. One of the most telling slides was that summary showing that 100 percent of those 14 trials surveyed tested visual acuity, and then it dropped down to 60 percent for field testing, and finally only two out of the 14 tested certain items. All of those tests take time.

Just to properly test the four things you mentioned—visual acuity, fields, color, and contrast—is a full day of testing. So cost is not the only problem; time and patience is also a problem. Obviously, it would be nice to measure all of these things, but it's just simply not practical because generally speaking in a clinical trial, your patients are very gracious to give you 1 day for all-day testing. But it's very difficult to get more than 1 day of testing, and you quickly run out of time.

**DR. KLEIN:** When doing our large population-based studies we wanted to add quality-of-life measures. But we found that some are quite time-consuming, and that brought up the same problems: what to put in, who to test it on, and whether the costs are justified by the benefits.

**DR. MURPHY:** Another practical issue is patient fatigue. Given that patients are willing to give you a whole day's time when you start doing all of this testing, by the end of the afternoon, they've really had a much more grueling day than the average older diabetic, or certainly the average patient with macular degeneration, is used to. The testing takes much more effort than they're used to expending in an average day.

**DR. FERRIS:** And the quality of the test results goes down commensurately.

In the ETDRS, we did all four of those tests, at some cost to our patients and our staff, and the one that was questionable was color vision. Some of our investigators said, "We might find changes in color vision over time, and they may be statistically significant." But do they pass the "so what" test? Does anybody really care if their color vision is down a little bit if their visual acuity is improved? If we're going to add on some quality-of-life tests, we're probably going to have to subtract some of the other tests that we're doing.

**MS. HAWKINS:** I was fascinated by the studies in which they compared six or seven different quality-of-life instruments. I wonder how the patients ever endured it.

**DR. BAROFSKY:** It's therapeutic completing quality-of-life assessments. Also, patients are willing and interested to undergo tests, especially if it's explained to them that this is an opportunity to have some input, drawing on their own points of view, into the treatment development process, which is particularly pertinent in clinical trials.

**DR. KLEIN:** There was an example of this in the Optic Neuritis Treatment Trial. Optic neuritis patients often complained about problems with visual function, although their visual acuity was 20/20 or 20/25. These complaints suggest the need for other tests like contrast sensitivity, which are appropriate to test visual function disturbances.

**MR. BUXTON:** Earlier, Dr. Legge said that there was room for improvement in the objective tests. Have we in fact covered some of the issues that you had in mind when you made that statement, or were you talking about something beyond the refinement of these standard areas?

**DR. LEGGE:** I think what I had in mind was going a little bit beyond the class of tests that Ron Klein talked about, to more function-oriented tests like reading performance rather than simply acuity. For example, we know that reading speed is not particularly highly correlated with acuity. There's obviously some correlation but it's not a particularly high correlation. The same may be true of other everyday tasks. You may be able to define certain mobility tasks or other real world functions to build a standardized test.

I would like to see performance measures that are studied and related to these sorts of clinical measures but that are a little bit closer to the kinds of everyday tasks that people do.

**DR. KLEIN:** In Stuart Fine's study of photocoagulation treatment of age-related macular degeneration, he has participants looking at clocks, and he has four or five basic tasks that relate to task-oriented visual function. Does this come closer to what you're suggesting, using visual acuity but using it in a more "functional" sense?

**DR. LEGGE:** I have a feeling that there's an infinity of different visual tasks we really do. But performance on many of them will be highly correlated, like reading a clock and reading money.

**MS. HAWKINS:** Everybody can read money. That's one of the tasks that didn't correlate!

**DR. LEGGE:** Actually, we have people come into our lab with low vision who can read reasonably well, but when we pay them they have to be told exactly what the bills are. American money is all green, and the print isn't all that great. In fact it's not that easy to read if you have very low vision.

**MS. HAWKINS:** In our study, patients had to have almost no light perception in order not to read money.

**DR. KLEIN:** When I was in residency training, Lawton Smith would do a visual field on patients who were confined to a hospital bed, almost comatose. He would take out different dollar bills to do a visual field. At that time, dollar bills were still of some value, and he'd move them around to get the patient to respond. That's how he tested the visual field in these very sick patients.

# General Health Measurement And Quality of Life in Visual Impairment

Ian McDowell, Ph.D.

Although measuring quality of life is fashionable, the concept is used inconsistently. Before considering general health measurement, therefore, it is important to establish a clear understanding of what quality-of-life measurement (QLM) is and how general health measurement relates to QLM.

This paper distinguishes three kinds of measurement related to quality of life: functional status measures, general health measures, and true quality-of-life measures. It discusses the uses and purposes of QLM and reviews some specific instruments currently used to measure general health as it relates to quality of life.

## Categories of Measurement

There is an unfortunate trend to rename morbidity or functional status measurements as quality-of-life scales. In reality, important differences remain between these kinds of measurement. Three major categories of measures include:

- *Functional status measures.* This category includes the familiar activities of daily living (ADL) scales, which are very often brief questionnaires covering commonly required activities such as climbing stairs, washing, and so forth. They are numerous and have existed for more than 30 years. The vast majority have not been evaluated very closely for reliability or validity.
- *General health status measures.* Unlike functional status indicators, general health status measurements take a multi-dimensional approach. They include the familiar ADL themes that characterize the functional status measurements but also usually cover psychological, physical, and often social functioning. In addition, they sometimes cover some quality-of-life dimensions such as the extent to which social support is available and the resources that the environment can offer the patient.

  General health status measures are generally more recent and better validated than functional status measures, and use more precise scoring procedures. They tend to be more suited to research applications than to brief clinical assessments.
- *Quality-of-life measures.* These scales cover the topics covered in the two categories described above, but also include personal, material, and environmental resources and opportunities for meeting needs. Generally defined as an overall evaluation of the subjective experience of life, quality-of-life scales are broader than general health measurements. For example, Hornquist (1982) proposed an approach to quality of life that considers the balance between a person's needs and resources in physical, mental, and social domains, and which considers the adequacy of material conditions, such as housing. He also stressed the relevance of evaluating "structural" factors that influence quality of life: access to political participation, the level of recognition of the needs of disabled people, social justice, and access to authorities. The main contrast between health status and quality-of-life measurements lies in the attention paid to environmental factors in the latter.

## Why Measure Quality of Life?

How far along this general continuum of measurements should we proceed for visually impaired persons? Which category of measurements should be used? To answer these questions, consider first the uses of QLM.

- QLM appears valuable for patients with life-threatening disease (e.g., cancer), to indicate

the quality of survival where therapies have adverse side effects. This may lead to a conflict with the physician's imperative to save life; quality of life is therefore distinguished from the value of life. Happily, these issues are not pertinent to visual impairments.

- QLM has been used to compare treatments that are equivalent in terms of other indices. For example, when indicators of morbidity, mortality, or return to work showed no difference between bypass graft surgery and medical therapy, QLM was used in an effort to distinguish between the two treatments.
- QLM may identify treatable psychological and other sequelae of illness.

Not only must we decide which of these applies to visual disorders, but also we must decide who judges quality of life. It could be the patient or medical staff. Or it could be an observer; quality of life might, for example, be judged by competent patients of the same age and diagnosis, following legal models (Pearlman and Jonsen, 1985). Alternatively, can quality of life be judged in objective terms? Could it be seen, for example, as the attainment of attributes highly valued in society? With changes in human rights legislation and with shifting attitudes toward professionals, the merits of these alternatives may be shifting as well.

## General Health Status Measures

General health scales are summarized in Tables 1 and 2. Although the list is incomplete, the tables can serve as an indication of the kinds of instruments that fall into this category. Of the instruments listed, only the Comprehensive Assessment and Referral Evaluation (CARE) includes significant coverage of vision (11 questions). Four of the other scales each have a single question on vision.

Two scales, the Multilevel Assessment Instrument and the Multi-Dimensional Functional Assessment Questionnaire, contain important components that could be considered as QLM, such as environmental conditions and need for services.

Table 1 indicates whether the measurement gives an ordinal scale or an interval scale. It also gives a rough indication of how long it takes to administer the tests. The shortest instrument, for example, contains 18 questions. Table 2 gives a brief outline of the numbers of questions related to each topic.

Of the measurements listed, several may be suited for use with patients with visual impairment:

- *The Functional Assessment Inventory* is an imaginative but clinical scale used for patient rehabilitation. It does cover something very close to the quality-of-life theme in its questions on personal and environmental resources. Included are questions on present skills, aspects of coping, cognitive abilities, wealth, and other resources.
- *The Sickness Impact Profile (SIP)* probably is suited for use with visual impairment, although it addresses a severity of impact that may be inappropriate.
- *The Multilevel Assessment Instrument (MAI)* is designed for elderly people in the community. It includes quality-of-life dimensions such as environmental characteristics and environmental resources. It has only one vision question, but it may be possible to build on this instrument.
- *The Multi-Dimensional Functional Assessment Questionnaire (MFAQ)* is designed for elderly people living at home. It is very broad ranging and could well be described as a quality-of-life scale. It covers the need for services, use of services, and other aspects of quality of life, and it is continuing to be thoroughly tested.
- *The CORE-CARE Instrument*, also designed for elderly people, covers many quality-of-life dimensions, such as the physical, mental, and social aspects of health and more. It includes a number of existing scales, and is built on relatively well-validated measurements. It has 11 questions on vision. Shorter versions are available.
- *The Quality of Well-Being Scale* focuses more on physical problems than visual or other sensory difficulties. It does, however, have a strong scoring method. It is the only test that provides a ratio scale, for example, and its inclusion of a prognostic dimension is ex-

**Table 1**
**Comparison of the Quality of General Health Measurements**

| Measurement | Scale | Number of Items | Application | Administered by (time) | Studies Using Method | Reliability | | Validity | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Testing Thoroughness | Results | Testing Thoroughness | Results |
| Arthritis Impact Measurement Scale (AIMS) (Meenan) | ordinal (Guttman) | 45 | clinical | self (15 min) | few | + + | + + | + + | + + |
| Physical and Mental Impairment of Function Evaluation (PAMIE) (Gurel) | ordinal | 77 | clinical | expert (10-15 min) | few | + | + + | + | + |
| Functional Assessment Inventory (Crewe & Athelstan) | ordinal | 40 | clinical | expert (5 min) | few | + + | + + | + + | + + |
| Nottingham Health Profile (Martini & Hunt) | interval | 38 | clinical, survey | self (<10 min) | several | + + | + + | + + | + + |
| Sickness Impact Profile (Bergner) | interval | 136 | survey | self (20-30 min) | many | + + + | + + + | + + | + + |
| Multilevel Assessment Instrument (Lawton) | ordinal | 147 | survey | interviewer (50 min) | few | + + | + + | + + | + + |
| Multi-Dimensional Functional Assessment Questionnaire (OARS) | ordinal | 120 (Part A) | clinical | interviewer (30 min) | several | + + | + + | + + | + + |
| CORE-CARE (Gurland) | ordinal | 329 | clinical | interviewer | several | + + | + + (full CARE not CORE CARE) | + + + | + + |
| Quality of Well-Being Scale (Bush) | ratio | 18 | research | interviewer (15 min) | several | + + | + + + | + | + + |

**Table 2**
**Comparison of the Content of General Health Measurements, Showing Numbers of Questions on Each Topic**

| Measurement | Mental Capacity | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Self-care | Mobility | Travel | Body movement | Home management | Medical condition | Senses | Cognition | Anxiety | Depression | Other | Work | Resources | Social support | Social interaction | Hobbies | Behavior problems | Communication | Other | Total |
| Arthritis Impact Measurement Scale (AIMS) (Meenan) | 3 | 7 | 2 | 6 | 6 | 5 | | | 6 | 6 | | | | | 4 | | | | | 4 |
| Physical and Mental Impairment of Function Evaluation (Gurel) | 15 | 2 | | | | 8 | 2 | 14 | 5 | 7 | | | | | 11 | | 12 | 1 | | 7 |
| Functional Assessment Inventory (Crewe & Athelstan) | 2 | 1 | | 5 | | 1 | 2 | 6 | | | 2 | 8 | 4 | 2 | 2 | | 1 | 4 | | 4 |
| Nottingham Health Profile (Martini & Hunt) | 1 | 5 | | 2 | | 11 | | | 3 | 7 | 1 | | | | 3 | | | | 5 | 3 |
| Sickness Impact Profile (Bergner) | 12 | 24 | 4 | 7 | 10 | 9 | | 10 | 3 | 3 | 1 | 9 | | | 20 | 8 | | 9 | 7 | 13 |
| Multilevel Assessment Instrument (Lawton) | 5 | 2 | 4 | 1 | 6 | 45 | 2 | 10 | 1 | 7 | 3 | | 25 | 7 | 27 | 2 | | | | 14 |
| Multi-Dimensional Functional Assessment Questionnaire (OARS) | 5 | 2 | 1 | 1 | 7 | 27 | 2 | 6 | 2 | 4 | 15 | 7 | 27 | 4 | 6 | 2 | 2 | | | 12 |
| CORE-CARE (Gurland) | | 27 | | 39 | | 107 | 25 | 19 | | 29 | | 7 | 34 | 34 | | | | | 8 | 32 |
| Quality of Well-Being Scale (Bush) | 1 | 3 | 2 | 4 | | 4 | | | | | | 4 | | | | | | | | 1 |

tremely interesting. While it is methodologically the most striking, it may not be as applicable to vision problems as some of the others.

## Conclusion

It appears that three of these instruments deserve further consideration: These are, in order, the CORE-CARE, MAI, and MFAQ instruments. Others of interest are possibly the SIP, and possibly the Quality of Well-Being Scale, with special reference to its scoring methods. Beware, however, the fashionable element in quality-of-life measurement: It is too often used as a buzz word to catch the eye of granting agencies, suggesting a certain nobility and humanity on the part of the researcher or therapist. Like democracy, it forms a rallying point for people of widely different persuasions; like democracy, the term may be used to create an effect while avoiding a precise definition.

## References

Hornquist JO. The concept of quality of life. *Scandinavian Journal of Social Medicine* 10:57-61, 1982.

Pearlman RA and Jonsen A. The use of quality-of-life considerations in medical decision-making. *Journal of the American Geriatric Society* 33:344-50, 1985.

# Discussion

**MR. BUXTON:** I think it might be appropriate to ask Marilyn Bergner whether the Sickness Impact Profile has been used with the visually impaired, and if so, to what effect.

**DR. BERGNER:** I don't have, nor do I have any way to get, a complete listing of studies in which the SIP has been used. I know of one person who is actually using the SIP to look at visual impairment and rehabilitation with the visually impaired. SIP seems to have items that can detect changes that would occur in the everyday functioning of patients.

One of the people very centrally involved in the development of the SIP, a psychologist named Ruth Bobbitt, is visually impaired. She was very concerned that we have items that might measure the effects of visual impairment. So although there are no items that ask about vision, there are items that ask about getting around in the dark, being able to get out by oneself, driving, reading, and performing fine handwork.

Hearing impairments raise the same issues. It's not the same disability but it has the same set of problems to solve in devising questions that have to do with everyday functioning. It is extremely difficult to devise such questions. We sent the early versions of the SIP to people who were in the field and asked, "Can you tell us if there are things that are missing?" People who were dealing with hearing impairment said that there weren't items that addressed the effects of that disability. And although we tried to develop such items, we really weren't able to. We just couldn't do it.

I, however, have used the SIP with a small group of hearing impaired who were having cochlear implants, and just recently Gene Nelson used it with a group of people with hearing impairment. Both of us found, to our surprise, that we were able to detect differences over time, and differences between people with different levels of hearing impairment.

One more comment about the length of time to administer the instruments: For the SIP, at least 25 or 30 minutes is the norm. We do get some extremely ill people who take a long time because they need to do it in bits and pieces. Or some people may be very slow in responding, but that's the extraordinary case.

**MS. HAWKINS:** With respect to people with vision impairment, we have to think a little about the method of administration; self-administration, unless you could do it with a tape, may pose difficulties.

**DR. BERGNER:** The SIP can be administered by an interviewer or be self-administered. It can be done either way.

**DR. BAROFSKY:** The tables in the paper we just heard from Ian McDowell give the impression that

some of the assessments may not be directly assessing sensory capacities or sensory functions. Many of them, however, do assess the impact of a deficit. For example, the Quality of Well-Being Scale has a lot of items that deal with the consequences of visual impairment.

**DR. PATRICK:** And two items that deal directly with visual impairment.

**DR. BERGNER:** There are direct questions about symptoms.

**DR. PATRICK:** There's one question about wearing glasses. I think it's important that we separate the different types of measures, too. The overview paper [Measuring and Valuing Health States Relating to Visual Impairment: A Review of Literature, Concepts, and Methods; see Appendix A] makes a good distinction between classification systems and questionnaires.

The Quality of Well-Being Scale is a classification system, like the Torrance measure mentioned in the review paper. People are classified to specific states and are given different levels of functional status and symptomatology for which utility weights have been assigned. That's different from a traditional questionnaire that may be administered with yes/no and true/false answers. The two different systems may give different results.

I don't know any measure that picks up rigidity of behavior due to an impairment or the cosmetic questions or the stigma questions. But as far as reading, driving, walking, watching TV, and pursuing hobbies and social activities are concerned, these are in many different general health measures, from the 100- to 200-item ones to the 20-item Medical Outcomes Study measure which covers a wide range of functional states with very few items.

**DR. KLEIN:** I have a question for Ian McDowell. You have a column in the table that's called validity. I may have missed it when you were presenting, but how did you define validity?

**DR. McDOWELL:** The idea of the pluses is the more the better, and the number of pluses ranges between zero and three. We essentially looked at the validity coefficients, the validity results, and ranked them and did a somewhat arbitrary assignment of pluses. We actually went through it several times blind to the previous results until the results became consistent.

It looks to me as if the three or four scales that we're contemplating are pretty equivalent, save the fact the SIP has had more reliability testing.

**DR. KLEIN:** The concept of reliability I understand, but tests of validity are still unclear. How do you evaluate the validity of a test? What are the validity measures that you would apply?

**DR. BERGNER:** One would like to think that the question items have something to do with health. Since we formulated those items by asking people who were sick how their behavior when they were sick differed from their behavior when they were well, we thought that the original group of items had at least some face validity.

We also took the items and submitted them to about 20 people who we thought were knowledgeable about illness, and asked, "Do these items cover the areas that you think are impacted by an illness? Do you think we're missing some things? Are we duplicating some things? Are we asking them in the wrong way?"

We got some sense of content validity in that way. We did have people criticize some of our items, and we did make changes. There also were issues about validating the scaling, but I won't go into that part of it.

We also wanted to know whether the SIP would detect changes over time. So we tested small groups of patients in whom there would be a relationship between everyday functioning and physiologic functioning. We tested hip replacement patients, hyperthyroid patients, and arthritic patients. We tested them over a period of time, and at times when the clinician told us that they expected that they would change, to see whether SIP changes and physiologic changes mirrored each other. We also looked at utilization of care among people who had higher scores versus those who had lower scores.

Also, the results of using these measurements give you some sense of validity. Some of my results are particularly striking and seem to relate to the disease process itself. For example, among hyperthyroid patients we found that there was a high level of sleep dysfunction. No one had ever seen this before. But it turned out that the literature on hyperthyroid conditions shows that there are EEG changes that should affect sleep, and so it was a plausible set of results.

**DR. KLEIN:** So if we construct questions for eye disease, we have to go through a great deal of testing and validation before they are used in a clinical trial. We are at an early stage.

Dr. Bernth-Peterson's talk illustrated that questions can be developed and used successfully. It just seems that in the eye health care field our approaches are primitive. Is that wrong?

**DR. BERGNER:** You're coming in at a time when there's a great deal of sophisticated work that has been done, not in vision, per se, but in other fields,

work that you can build on. If you were doing this 15 years ago, you wouldn't have had this much. And you would have really been in a primitive stage; everybody was in a primitive stage then.

**DR. FERRIS:** That's why we didn't do it 5 or 10 years ago. We thought about it and threw our hands up because there was no base to work with.

**MR. BUXTON:** Do those people who are familiar with quality-of-life instruments feel that there are any important ones that are not yet being discussed?

**DR. WARE:** I think the eye health care field has several good measures of visual functioning. So I think you're coming from a position of strength. In that regard, I think the real challenge to your research agenda lies between the more specific and generic quality of life measures. That's what Donald Patrick is going to talk about tomorrow.

*I think the eye health care field has several good measures of visual functioning. So I think you're coming from a position of strength.*

**MR. McMANUS:** One of the policy questions that we would like your evaluation and input on is this: "Is it feasible to use some of these generic scales?" If so, we would be able to compare results for visual impairment with those for other medical conditions.

**DR. PATRICK:** Yes. You could learn from something like the Arthritis Impact Measurement Scale (AIMS) which includes reading, driving, ambulation, TV watching, hobbies, and sports, all as they relate to arthritis. In working with the pharmaceutical industry over the last 2 years, I've found that people are doing trials with respiratory patients using the AIMS, which shows the generic quality of the scale.

And so my personal preference is that you don't create another instrument with generic concepts that are contained in generic instruments, and give it to condition-specific populations and create a so-called specific instrument. If you do, you then have the problem that you cannot compare the measure with other types of measures.

If you need the generic instrument, if you're going to do the cross-condition comparison, that does not preclude a condition-specific supplement that is tied into the generic instrument with the same scoring system. This has been done with the SIP for conditions like head injury by adding items and going through elaborate testing without making much difference in the measurement capabilities. You still have the generic instrument for comparison.

**MS. HAWKINS:** One problem with modifying instruments is that it's very hard to find out about all the instruments that are available, get copies of all the questionnaires, and get copies of all the administration and coding manuals. Even at this meeting, I have yet to see a list of things that have been done, where to get up-to-date copies, where to get information.

Also, in longitudinal studies (which clinical trials are), there's the problem of patients memorizing instruments, such as their visual acuity charts. The same issue arises with these questionnaires. Are there multiple versions of any of these questionnaires? Are there instruments that are interchangeable?

**DR. WARE:** For most study designs, it's not necessary.

**MS. HAWKINS:** It all depends on how frequently you're testing your patient, how long a trial you're doing, how fast the patient's visual function is changing, and so forth. In many long-term chronic illnesses, multiple versions may not be necessary, but I would suspect there are shorter-term studies.

**DR. BERGNER:** Some of the instruments may have tested that.

**DR. McDOWELL:** It's possible also to overstate the need for alternate versions because there is an important idea in the literature that suggests that you perhaps should show the patient their previous responses, rather than treating it as a knowledge test. You could say, "This is what you said your pain was like 2 weeks ago. Now, in the light of that, has it gotten better or worse?"

**DR. FERRIS:** I can't help but make one last comment about the visual acuity exam. It isn't just that it was quick and easy. The reason that it's lasted so long, I believe, is that ophthalmologists and others have felt that it does have something to do with function. I would be interested in further discussions as to whether visual acuity may indeed be correlated to some of the scales that already exist.

# Disease-Specific Versus General Health State Measures

Donald L. Patrick, Ph.D.

Increasing knowledge and use of health status and quality-of-life assessment in clinical research has spawned a debate over the relative merit of generic versus disease-specific measures. Generic health status measures are those that purport to be broadly applicable across types and severities of disease, across different medical treatments or health interventions, and across demographic and cultural subgroups. These measures are designed to summarize a spectrum of the concepts of health or quality of life that apply to many different impairments, illnesses, patients, and populations.

Disease-specific measures are those designed to assess specific diagnostic groups or patient populations, often with the goal of measuring responsiveness or clinically important changes. These are changes that clinicians and patients think are discernible and important, have been detected with an intervention of known efficacy (Guyatt et al., 1987), or are related to well-established physical or physiological measures, e.g., visual acuity or field for eye patients.

This paper reviews the characteristics and uses of generic and disease-specific measures and compares their advantages and disadvantages with regard to reliability, validity, responsiveness, interpretation, and generalizability.

## Generic Measures

Health status measures over the last 2 decades have gradually been broadened to include a wide number of concepts including that of quality of life. Generic measures of health status summarize concepts of health-related quality of life.

Health-related quality of life (HRQOL) is a broad term covering five categories of concepts:

- Duration of life
- Impairments
- Functional states
- Perceptions
- Social opportunities.

These are health-related to the extent they are influenced by disease, injury, treatment, or policy (Patrick and Erickson, 1988).

Generic measures may provide operational definitions of several concepts summarized by a single index value or in a profile of interrelated scores. Table 1 shows the major concepts contained in six well-known generic health status measures.

Measures such as the Quality of Well-Being Scale (QWB) (Bush, 1984), the Torrance Classification (Torrance, 1986), and the Rosser Scale (Rosser, 1983, 1988) combine duration of life with specific dimensions of impairment and physical, psychological, and social function. These measures yield a single index value, called quality-adjusted life-years (QALYs), that can be used to compare the cost per QALY gained from different health interventions.

Other generic measures provide a profile of scores for different components of health-related quality of life. One major example, the Sickness Impact Profile (SIP) (Bergner et al., 1981), assesses sickness-related dysfunction in 12 different categories, producing a score for each category. Various categories may be aggregated into a physical dimension score, a psychosocial dimension score, and an overall score with independent categories of work, eating, sleep and rest, home management, and recreation and pastimes.

Similarly, Part I of the Nottingham Health Profile (NHP) (McEwen, 1988) contains 38 items that cover six domains of experience yielding individual scores for each; Part II contains perceived problems in seven areas of daily life. Unlike the SIP, however, the NHP does not yield an overall score.

## Table 1
## Major Concepts of Health-Related Quality of Life Contained in Selected Measures

| Concept Dimension | Generic Measures | | | | | |
|---|---|---|---|---|---|---|
| | Health Insurance Experiment[1] and MOS[a] Short Form[2] | McMaster Index[3] | Nottingham Health Profile[4] | Quality of Well-Being Scale[5] | Rosser Classification[6] | Sickness Impact Profile[7] |
| Opportunity[b] | | + | + | | | |
| Perceptions[c] | + | + | + | | | |
| Social Function[d] | + | + | + | + | | + |
| Psychological Function[e] | + | + | + | | + | + |
| Physical/Role Function/Fitness | + | + | + | + | + | + |
| Impairment[f] | + | + | | + | | + |
| Death | | | | + | + | |

[a]Medical Outcomes Study
[b]Includes handicap and resilience
[c]Includes general perceptions and life satisfaction
[d]Includes social contact, integration, and intimacy
[e]Includes affective and cognitive status
[f]Includes symptoms, signs, self-reported of disease or condition, physiologic measures, tissue alterations, and diagnoses


1. Chambers LW. The McMaster Health Index Questionnaire: An update. *In:* Walker SR and Rosser Rm. Quality of life: Assessment and application. Lancaster, England: MTP Press, 1988; 113.
2. Brook RH, Ware JE, Davies-Avery A et al. Overview of adult health status measures fielded in Rand's Health Insurance Study. *Medical Care* 1979; 17 (7 suppl).
3. Stewart AL, Hays RD, Ware JE. The MOS Short-Form General Health Survey: Reliability and validity in a patient population. *Medical Care* 1988; 26:724.
4. Deyo RA, Inui TS, Leininger J et al. Physical and psychosocial function in rheumatoid arthritis. *Arch Intern Med* 1982; 142:879
5. Vanden Burg MJ. Measuring the quality of life of patients with angina. *In:* Walker SR and Rosser RM. Quality of Life: Assessment and application. Lancaster, England: MTP Press, 1988; 267.
6. Deyo RA and Hiehl AK. Measuring physical and psychosocial function in patients with low-back pain. *Spine* 1983; 8:635.
7. Follick MJ, Smith TW, Ahern DK. The SIP: A global measure of disability in chronic low-back pain. *Pain* 1985; 21:67.

The 59-item McMaster Health Index Questionnaire (Chambers, 1988) yields indexes for physical, emotional, and social function. Measures developed originally by The Rand Corporation—the 108-item Health Insurance Study battery (Brook et al., 1979) and the 20-item Medical Outcomes Study short-form generic measure (Stewart et al., 1988)—cover a wide spectrum of health concepts for use in general populations.

All these generic measures have been tested extensively with different patient populations. For example, the SIP has been used in the United States and abroad with a large number of different populations, including subjects having the following conditions:

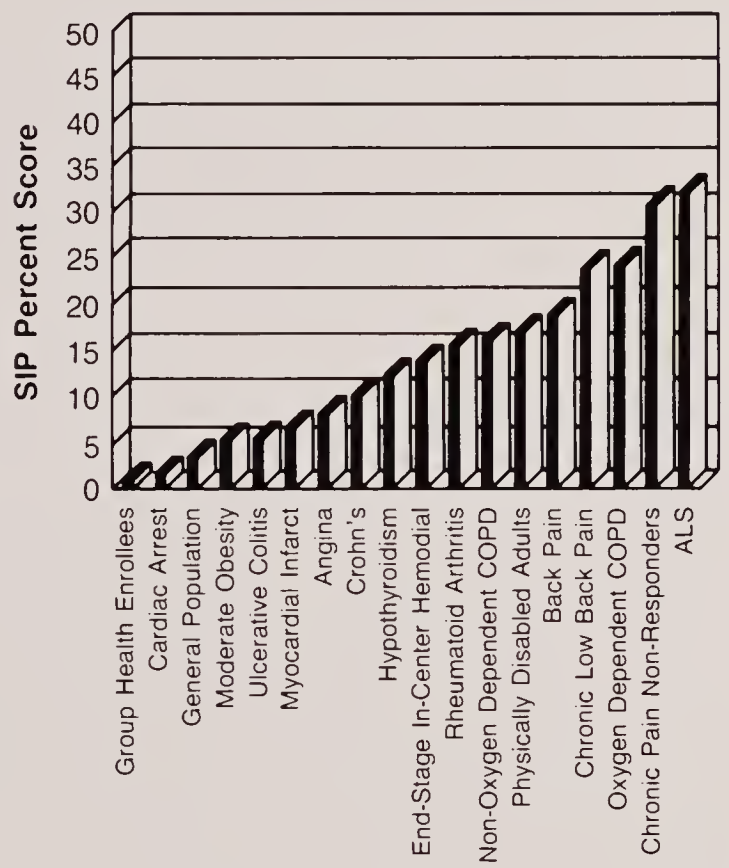
- rheumatoid arthritis (Deyo et al., 1982)
- angina (Vandenburg et al., 1988)
- back pain (Deyo and Diehl, 1983; Follick et al., 1985)
- cancer (Sugarbaker et al., 1982)
- obesity (Sullivan et al., in press)
- end-stage renal disease (Hart and Evans, 1987)
- myocardial infarction (Ott et al., 1983; Bergner et al., 1985)
- benign chronic pain (Augustinsson et al., 1986)
- chronic obstructive pulmonary disease (COPD) (McSweeny et al., 1982; Bergner et al., 1988)
- cardiac arrest (Bergner et al., 1985; Bergner et al., 1984)
- amyotrophic lateral sclerosis (Fink et al., 1985)
- ulcerative colitis (Drossman et al., 1989)
- Crohn's disease (Drossman et al., 1989)
- hyperthyroidism (Rockey and Griep, 1980)
- physical disability (Patrick and Peach, 1989).

In addition, several populations of healthy adults have been investigated (Bergner et al., 1985; Patrick and Peach, 1989; Gilson et al., 1979).

Figure 1 shows overall SIP percent scores for these different populations, ranging from minimal to maximal dysfunction (e.g., low to high scores). Most of the scores shown in Figure 1 are not standardized for age or sex, although all studies reported here were conducted with adults and most obtained data from both men and women.

Overall SIP scores for general populations of primarily well adults or those who have survived myocardial infarction or cardiac arrest are at the low end of the dysfunction continuum. Scores for patients with chronic, debilitating illnesses such as rheumatoid arthritis, chronic pain, oxygen-dependent COPD, or amyotrophic lateral sclerosis are considerably higher. When used to compare these cross-sectional samples, the SIP appears to discriminate in expected directions among different populations and different degrees of disease severity.

**Figure 1**
**Overall Sickness Impact Profile (SIP) Scores for Different Disease Conditions or Population Groups**




## Disease-Specific Measures

Disease-specific measures have been used extensively over several decades. Early examples include the Karnofsky Performance Status Scale for cancer (Karnofsky et al., 1948; Karnofsky and Burchenal, 1949), the American Rheumatism Association (ARA) functional classification for ar-

thritis (Steinbrocker et al., 1949) and the New York Heart Association Functional Classification (The Criteria Committee of the New York Heart Association, 1964).

Measures like these were developed to meet the need for rapid classification of patients. However, their sensitivity to small but clinically important change is probably limited. The ARA classification, for example, may detect large changes such as those following hip replacement but not smaller changes following drug therapy judged successful by other criteria (Deyo, 1984).

The popularity of modern disease-specific measures arises primarily from the need of clinical trials and practitioners to use scales that are most responsive to clinical changes that occur over time (Guyatt et al., 1987; Deyo, 1984; Krischner and Guyatt, 1985). Both discriminating improved from unimproved patients and accurately quantifying minimally important changes are particularly important measurement objectives for clinical research and clinical practice.

Although items concerning breathlessness, pain, and functional status are often part of generic measures, disease-specific measures tend to use different wording for items and instructions tailored to the diagnostic group, condition, or population under investigation. Items in arthritis-specific measures, for example, may include the phrase "due to my arthritis." Cancer-specific measures include items on nausea and vomiting, common side effects of chemotherapy. Disease-specific measures may also have greater detail or more items concerning specific functions, e.g., visual function.

## Models for Using Generic and Disease-Specific Measures

Four different models seem relevant in examining the use of generic and disease-specific measures to date.

The first approach is to include both generic and disease-specific measures in the same investigation even though the concepts covered by the different instruments may overlap substantially. Thus, one might include a generic measure containing visual impairment items as well as a vision-specific functional scale.

The second approach is to compare a generic instrument and a generic instrument modified for the specific population of interest. Items pertaining to visual impairment, disability, or handicap might be added to an existing instrument or items might be selected or reworded specifically for eye patients.

The third approach is to use a generic health status instrument with a condition-specific supplement. This is similar to the first approach except that the condition-specific measure is constructed with a different conceptual basis and minimal overlap with the generic measure. The intention is not to measure the same concepts as a generic measure with specific reference to a medical condition, but to capture the additional, specific concerns of patients that are not contained in generic measures. For example, the generic instrument might assess functional status, while the condition-specific instrument would address the specific handicaps or disadvantages experienced by visually impaired persons.

The fourth approach is to administer health status batteries, which are collections of specific measures that are scored independently and reported as individual scores (Patrick and Ericksen, 1988). Although generic instruments may be included in health status batteries, collections of specific measures are more often used. Batteries are common in clinical trials and epidemiological investigations, where entire scales, subscales, or individual items from the best available instruments are administered and effects are tested for each measure in the battery.

## Advantages and Disadvantages Of Generic and Disease-Specific Measures

The relative advantages and disadvantages of generic and disease-specific measures depend primarily on the specific objective of measurement and its application in clinical research or policy analysis. HRQOL measures may be used to discriminate among residents at a point in time, to predict future outcomes or events, and to measure changes over time (Krischner and Guyatt, 1985). Technical requirements vary considerably according to the objective of measurement and the specific environment of the application.

In a given study, there are often competing measurement objectives, such as discrimination, prediction, and change analysis. In such cases, measurement priorities are necessary. In addition, possible technical requirements for a given application include validity, reliability, responsiveness, ease of analysis, and generalizability.

*The relative advantages and disadvantages of generic and disease-specific measures depend primarily on the specific objective of measurement and its application in clinical research or policy analysis.*

### Validity

Content validity depends on the extent to which an empirical measurement reflects the full domain relevant to the particular measurement situation (Carmines and Zeller, 1979). The assessment of health status for persons with visual impairment may well include the full range of concepts and dimensions that constitute health-related quality of life. Generic measures used in condition- or population-specific applications may have low content validity because they contain items of little or no relevance to study participants. For example, items concerning incontinence or eating behavior may not concern patients with visual problems. In addition, generic measures may miss concerns of particular interest to study participants, because the process of item selection did not include patients with this condition or their relatives and clinicians.

By contrast, the focus of disease-specific measures gives them high content validity for clinical investigations of special populations. For example, items specifically concerning vision and visual deficit are of major concern to those with visual impairment. Specifying in advance which items are relevant, however, is not always possible in an investigation, particularly when testing a new treatment or attempting to affect a broader aspect of quality of life such as social function or satisfaction.

Construct validity of a health status measure depends on two related concepts: (a) convergent evidence or expected correlations among different measures such as age, medical conditions, and use of services, and (b) discriminant evidence or predicted differences between measures such as functional status and personality (Kaplan et al.. 1976). Advance specification of expected differences according to some theory of state or trait relationships is necessary to establish this type of validity.

The construct validity of some generic health status measures has been established (Brook et al., 1979; Stewart et al., 1988; Bergner et al., 1981; Kaplan et al., 1976) although no general theory of expected differences has emerged. The construct validity of disease-specific measures is less well established, despite recent attempts to correlate disease-specific measures with patient and clinician global ratings and with disease-activity indexes (Guyatt et al., 1986).

### Reliability

Reliability is the extent to which a measuring procedure yields the same results on independently repeated trials under the same conditions (Carmines and Zeller, 1979). Test-retest reliability is measured using the Pearson's correlation or Kappa statistic (intraclass correlation), depending on the measurement characteristic of interest. Internal consistency reliability is usually assessed using Cronbach's alpha (Cronbach, 1951). Test-retest reliability is most appropriate for assessing static states or traits; the assessment of dynamic states is problematic because it is difficult to distinguish between random variation and true change.

Global measures created from generic health status measures often have high test-retest and internal consistency reliability, although individual items or subscales can have markedly reduced reliability coefficients (Bergner et al., 1981). Studies comparing the test-retest reliability of disease-specific measures with that of generic measures have not yet been published.

### Responsiveness

The ability of a measure to detect changes and the analysis of change scores or effect size are

important considerations for all serial applications of health status measures. Including items sensitive to change is critical to such assessments. Responsiveness of health status measures in clinical research has been assessed using the relative efficiency statistic, a ratio of paired *t* statistics (Liang et al., 1985); correlation of scale changes with other measures (Deyo and Centor, 1986); receiver operating characteristic curves (Deyo and Centor, 1986); and a responsiveness statistic, a ratio of minimal clinically important difference to variability in stable subjects (Guyatt et al., 1987). These approaches use similar data in calculating a statistic of responsiveness, although their mathematical relationships have not yet been investigated.

Disease-specific measures with items selected to assess particular concerns or worded to attribute change to the condition of interest, e.g., "due to my vision," may be particularly sensitive to within-subject changes and thus more responsive than generic measures that contain items unrelated to change. A responsive measure, however, may not demonstrate high test-retest or internal consistency reliability (Guyatt et al., 1987). In this case, a tradeoff must be made between responsiveness and reliability.

*The ability of a measure to detect changes and the analysis of change scores or effect size are important considerations for all serial applications of health status measures. Including items sensitive to change is critical to such assessments.*

### *Ease of Analysis*

Analyzing and interpreting changes in health status measures are problems for all longitudinal studies such as clinical trials. Changes in physiologic measures, such as blood pressure or cholesterol level, or in physical measurements, such as visual acuity, may be interpreted in terms of prognostic implications and well-established or agreed cut-off points.

Changes in generic health measures are more difficult to interpret, although even small changes in portions of such measures may be quite useful; for example, changes in physical mobility or self-care are meaningful in disabled populations. Changes in scores on the most general measures, such as health perceptions or global physical and psychosocial dimension scores, can be even more difficult to interpret.

The net changes observed may reflect a large number of different transitions or combinations of transitions within the population. Single-score or aggregated measures can make it difficult to identify which items or components are responsible for the change. Net changes also must be distinguished from random or systematic changes (learning effects, rumination) that may occur independently of an intervention. Although changes in these scores may reflect sensitive effects, the relative magnitude of the change may be difficult to assess. For example, is a five-point difference more meaningful than a three-point difference?

Changes in disease-specific measures may be easier to interpret because they are more specific or more closely associated with changes in clinical measures of disease activity such as blood pressure or joint inflammation. Clinician or patient assessments of improvement, which are common measures of change or effects, may be more closely associated with changes in disease-specific measures than with those in generic health status measures (MacKenzie et al., 1986).

Changes observed in studies using the battery approach may not be easy to interpret. Different results in different measures may make interpretation difficult, particularly when multiple comparisons or statistical tests are necessary (Guyatt et al., 1986). In studies using batteries, one measure may need to be chosen at the outset as the primary health status outcome to guide interpretation of changes.

### *Generalizability*

Comparability across different diseases, conditions, populations, or even investigations re-

quires a generic approach to health status assessment. The development of standardized, generic measures has permitted comparisons across different patient populations for clinical research (see Figure 1) and across different interventions in the evaluation of health outcomes for the purpose of resource allocation (Torrance, 1986; Kaplan and Bush, 1982).

Generic measures provide a common denominator or common unit of outcome by which to judge the relative severity of health outcomes and the relative effectiveness of health interventions. Many interventions also affect outcomes that are not condition-specific. For example, prevention programs to reduce cigarette smoking also reduce the incidence, prevalence, and adverse outcome of heart disease, cancer, and pulmonary disease. Investigators may not anticipate all treatment effects assessed with generic measures.

Disease-specific measures developed on special populations are not intended for general application. The issue of generalizability may still arise, however, in relation to the disease, condition, or population under investigation. Arthritis-specific measures have been applied across a large number of investigations, some involving patients with osteoarthritis and others patients with rheumatoid arthritis. Disease-specific measures need to be "general" enough for application to different sub-populations to permit comparisons between different organ-specific diseases.

The cross-cultural content of health status measures is relevant to both types of measures. Cross-cultural studies of preference weighting using the SIP (Patrick et al., 1985) and the Nottingham Health Profile (Hunt and Wiklind, 1987) suggest a universal ranking or rating of health states; for instance, it is better to walk without limitations than to walk with limitations. QWB preference weights obtained from the general population and from patients with rheumatoid arthritis were also similar (Balaban et al., 1986) and produced comparable results in cost-effectiveness analyses (Thompson et al., 1988).

Although rankings of health states may hold for different cultures and different disease-specific populations, the meaning of health status may well vary among these different populations. In one sense, disease-specific measures are translations of generic health status concepts into the cultural categories, vocabularies, and perspectives of different groups. For example, patients with visual impairment are a subculture; they communicate and evaluate the effectiveness of treatments in somewhat unique terms. Cumulative experience with generic health status measures has prompted the development of standardized instructions, interviewer-training procedures, computer scoring algorithms, and analytic strategies. Use of the same instrument by many different investigators with different populations requires standardized administration. This standardization arises, in part, because of the length and complexity of many generic measures with comprehensive components. Efficiency and effectiveness in administration requires detailed knowledge of the instrument and its uses. Generic measures such as the QWB and SIP require a major commitment to interviewer training, monitoring, and evaluation (Read et al., 1987), as well as resources for manipulating the data.

Although standardization is a marked advantage of generic measures, the length and complexity of some measures are not. Complex questionnaires and scoring algorithms may discourage the use of measures by interested investigators. Thus, short-form generic health status measures are beginning to appear in the literature for self-administration in clinical settings (Nelson et al., 1987; Jette et al., 1986) and in general population surveys (Stewart et al., 1988). Measures for use in clinical practice must meet the most strenuous criteria of minimum length, maximum validity, reliability, responsiveness, and low cost.

## Conclusion

Disease-specific measures may enjoy the advantages of ease in administration, acceptability, cost, and simplicity in scoring precisely because they are designed for application in clinical settings. Time constraints usually require that these measures be short, self-administered, highly acceptable to patients, and simple to score. The

tradeoff between short, simple measures and more comprehensive, complex measures in terms of reliability, validity, and responsiveness remains to be assessed for many generic and disease-specific instruments.

*Continued use of generic measures is necessary for comparing benefits of different health interventions and allocating resources.*

Both generic and disease-specific assessments are useful for clinical trials. Selection of particular measures depends on the objectives of measurement and environment of the application. In designing an optimal strategy, investigators need to assign priorities to measurement objectives.

Clinical and policy objectives are modified by the resources of the investigator, site of data collection, health and endurance of respondents, and constraints on time. Project aims, methodological concerns, and practical considerations often will determine whether generic or specific measures are used in an investigation.

No single general purpose measure is likely to meet all the needs of investigators and specific populations. Patients with different medical conditions have different concerns or place different emphasis on generic concepts of health. Rather than develop disease-specific measures that incorporate generic concepts of health-related quality of life, the preferred strategy is to use standardized, generic instruments with disease-specific supplements. Few investigations can afford overlapping measures, which also may have the disadvantage of irritating respondents because of repetitious questions.

Generic measures permit the comparison of different populations and different programs, a most important objective for policy analysis and decision-making. Continued use of generic measures is necessary for comparing benefits of different health interventions and allocating resources. Cumulative knowledge of health and quality-of-life outcomes using generic measures will establish the relative burden of different diseases and the relative merit of different interventions.

By contrast, instruments specific to different diseases, conditions, and populations are critical for identifying important concerns of patients with particular conditions and for measuring small, clinically important changes resulting from specific treatments. Experience with disease-specific measures to date indicates their usefulness in discriminating among different conditions and in assessing changes. Rapid development of such instruments is to be expected for conditions and populations where few specific measures now exist, such as patients with visual impairment. Visual impairment is a common condition in which specific patient concerns may be important outcomes of treatment.

Studies are needed to compare the validity, reliability, and responsiveness of generic and disease-specific measures with the same populations. These comparisons might well occur in clinical trials or cohort studies of specific populations. Particular attention should be paid to measuring change across two or more time intervals. Knowledge of transitions between different levels of health-related quality of life is limited. Evidence from longitudinal studies will help in understanding treatment outcomes, interpreting change scores, and estimating population prognoses.

## References

Augustinsson LE, Sullivan L, Sullivan M. Physical, psychologic and social function in chronic pain patients after epidural spinal electrical stimulation. *Spine* 11:111, 1986.

Balaban D, Sagi PC, Goldfarb NI et al. Weights for scoring the Quality of Well-Being instrument among rheumatoid arthritics: A comparison to general population weights. *Medical Care* 24:973, 1986.

Bergner LH, Hallstrom AP, Bergner M et al. Health status of survivors of cardiac arrest and of myocardial infarction controls. *American Journal Public Health* 75:1321, 1985.

Bergner L, Hallstrom A, Eisenberg M. Health status of survivors of out-of-hospital cardiac arrest six months later. *American Journal Public Health* 73:508, 1984.

Bergner M, Hudson LD, Conrad DA et al. The cost and efficacy of home care for patients with chronic lung disease. *Medical Care* 26:566, 1988.

Bergner MB, Bobbitt RA, Carter WB et al. The SIP: Development and final revision of a health status measure. *Medical Care* 19:787, 1981.

Brook RH, Ware JE, Davies-Avery A et al. Overview of adult health status measures fielded in Rand's Health Insurance Study. *Medical Care* 17(7 suppl), 1979.

Bush JW. General health policy model/Quality of Well-Being (QWB) scale. *In:* Wenger NK, Mattson ME, Furber CD, et al. (eds), Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies. New York: LeJacq Publishing Inc., 1984.

Carmines EG and Zeller RA. *Reliability and Validity Assessment.* Beverly Hills, CA: Sage Publications, 1979.

Chambers LW. The McMaster Health Index Questionnaire: An update. *In*: Walker SR and Rosser RM, Quality of Life: Assessment and Application. Lancaster, England: MTP Press, 1988.

Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297, 1951.

Deyo RA. Measuring functional outcomes in therapeutic trials for chronic disease. *Controlled Clinical Trials* 5:223, 1984.

Deyo RA and Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. *Journal Chronic Diseases* 11:897, 1986.

Deyo RA and Diehl AK. Measuring physical and psychosocial function in patients with low-back pain. *Spine* 8: 635, 1983.

Deyo RA, Inui TS, Leininger J et al. Physical and psychosocial function in rheumatoid arthritis. *Archives Internal Medicine* 142: 879, 1982.

Drossman DA, Patrick DL, Mitchell CM et al. Health-related quality of life in inflammatory bowel disease: Functional status and patient worries and concerns. *Digestive Diseases and Sciences* 1379-1386, 1989.

Fink AL. Social dependency and self-care agency: A descriptive-correlational study of ALS patients. Master of Nursing thesis, University of Washington, 1985.

Follick MJ, Smith TW, Ahern DK. The SIP: A global measure of disability in chronic low-back pain. *Pain* 21:67, 1985.

Gilson BS, Bergner M, Bobbitt R et al. *The Sickness Impact Profile: Final development and testing*. Department of Health Services, School of Public Health, University of Washington, 1979.

Guyatt GH, Mitchell A, Irvine EJ et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology* 96(3): 804-810, 1989.

Guyatt GH, Bombardier C, Tugwell PX et al. Measuring disease-specific quality of life in clinical trials. *Canadian Medical Association Journal* 134(8): 889-895, 1986.

Guyat GH, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 2:171, 1987.

Hart LG and Evans RW. The functional status of ESRD patients as measured by the SIP. *Journal of Chronic Diseases* 40(suppl): 117S, 1987.

Hunt SM and Wiklind I. Cross-cultural variation in the weighting of health statements: A comparison of English and Swedish valuations. *Health Policy* 8:227, 1987.

Jette AM, Davies AR, Cleary PD et al. The Functional Status Questionnaire: Reliability and validity when used in primary care. *Journal of General Internal Medicine* 1:143, 1986.

Kaplan RM and Bush JW. Health-related quality-of-life measurement for evaluation research and policy analysis. *Health Psychology* 1:61, 1982.

Kaplan RM, Bush JW, Berry CC. Health status: Types of validity and the Index of Well-Being. *Health Services Research* 11:478, 1976.

Karnofsky DA and Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. *In*: MacLeod CM, ed, Evaluation of Chemotherapeutic Agents. New York: Columbia University Press, 1949.

Karnofsky DA, Abelmann WH, Craver LF et al. The use of nitrogen mustards in the palliative treatment of cancer. *Cancer* 1:634, 1948.

Krischner B and Guyatt G. A methodological framework for assessing health indices. *Journal of Chronic Diseases* 38:27, 1985.

Liang MH, Larson MG, Cullen KE et al. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheumatism* 28: 542, 1985.

MacKenzie CR, Charlson ME, Digioia D et al. Can the Sickness Impact Profile measure change? An example of scale assessment. *Journal of Chronic Diseases* 39:429, 1986.

McEwen J. The Nottingham Health Profile. *In*: Walker SR, Rosser RM, Quality of Life: Assessment and Application. Lancaster, England: MTP Press, 1988.

McSweeny J, Grant I, Heaton RK et al. Life quality of patients with chronic obstructive pulmonary disease. *Archives of Internal Medicine* 142:473, 1982.

Nelson E, Wasson J, Kirk J et al. Assessment of function in routine clinical practice: Description of the COOP Chart method and preliminary findings. *Journal of Chronic Diseases* 40(suppl):55S, 1987.

Ott CR, Sivarajan ES, Newton KM et al. A controlled randomized study of early cardiac rehabilitation: The SIP as an assessment tool. *Heart and Lung* 12:162, 1983.

Patrick DL, Sittampalam Y, Somerville S et al. A cross-cultural comparison of health status values. *American Journal of Public Health* 75:1402, 1985.

Patrick DL and Peach H. A Sociomedical Approach to Disablement. Oxford: Oxford University Press, 1989.

Patrick DL and Erickson P. Assessing health-related quality of life for clinical decision making. *In*: Walker SR and Rosser RM, Quality of Life: Assessment and Application. Lancaster, England: MTP Press, 1988.

Read LJ, Quinn RJ, Hoefer MA. Measuring overall health: An evaluation of three important approaches. *Journal of Chronic Diseases* 40(suppl): 7S, 1987.

Rockey PH and Griep RJ. Behavioral dysfunction in hyperthyroidism: Improvement with treatment. *Archives of Internal Medicine* 140:1194, 1980.

Rosser RM. A health index and output measure. *In*: Walter SR, Rosser RM, (eds), Quality of Life: Assessment and Application. Lancaster, England: MTP Press, 1988.

Rosser R. Issues in measurement in the design of health indicators: A review. In: Culyer AJ (ed), *Health Indicators.* Oxford, England: Office of Health Economics, 1983.

Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. *Journal of the American Medical Association* 140:659, 1949.

Stewart AL, Hays RD, Ware JE. The MOS Short-Form General Health Survey: Reliability and validity in a patient population. *Medical Care* 26:724, 1988.

Sugarbaker PH, Barofsky I, Rosenberg SA, Gianola FJ. Quality-of-life assessment of patients in extremity sarcoma clinical trials. *Surgery* 91:17, 1982.

Sullivan ME, Sullivan LGM, Kral JG. Quality-of-life assessment in obesity: Physical, psychological and social function. *Gastroenterology Clinics of North America* 16(3): 433-442, 1987.

The Criteria Committee of the New York Heart Association, Inc. Disease of the Heart and Blood Vessels: Nomenclature and Criteria for Diagnosis 6th ed. Boston: Little, Brown, 1964.

Thompson MS, Read LJ, Hutchings HC et al. The cost effectiveness of auranofin: Results of a randomized clinical trial. *Journal of Rheumatology* 15:35, 1988.

Torrance GW. Measurement of health state utilities for economic appraisal. *Health Economics* 5:1, 1986.

VandenBurg MJ. Measuring the quality of life of patients with angina. *In*: Walker SR and Rosser RM, Quality of Life: Assessment and Application. Lancaster, England: MTP Press, 1988.

# Discussion

**DR. WARE:** Improving the specificity of measurement to make it more sensitive to the particular disease and the likely effects of treatment is probably one of the more important questions that this group is going to consider in regard to incorporating general health assessment in clinical trials.

Dr. Patrick has pointed out some very important distinctions, some subtle, in the way that an operational definition might be constructed. And yet I think that the results are going to shift dramatically with some of these subtle distinctions if visual impairment is like any of the other conditions that we have had experience with.

I am concerned that, for example, a general health assessment is not going to be sensitive to the dysfunction and stress that's associated with visual impairment. And I base that on a concern that people don't consider visual impairment a health problem.

We've seen examples of that in using generic measures. Take social activity, for example, such as frequency of getting together with friends and relatives. We've learned that when we ask those kinds of questions, they are not at all sensitive to health impact, the reasons being that variations in social activity, if not attributed specifically to health, are due to many things other than health. Therefore, you are really measuring 10 percent of what you are interested in and 90 percent of what you are not interested in. Measures like the SIP that specifically ask for attributions to health problems for variations in social activity are much more sensitive to the impact of disease and treatment.

If you carry that a step further, to whether an attribution is to a particular health problem as opposed to another health problem, you might get another major increment in validity or sensitivity. In other words, is the difficulty walking up and down stairs due to your arthritis or are the limitations on your social activities due to visual impairment?

I think this is so important. Yesterday someone suggested essentially eliminating the concepts that relate to activities of daily living, because a preliminary result suggested there wasn't any action there. I wonder what that study would have concluded if it had asked about limitations in activities of daily living due to visual impairment. It's kind of unthinkable that there wouldn't be any impact in 15 ADL items due to visual impairment. And I wonder if just the wrong question was asked.

**DR. PATRICK:** I'll give you the analogy of age. It's inconceivable to me that if you are very old and your bones are creaking, that walking up and down stairs is not a problem. But if you ask older persons SIP items like that, they'll say that's not due to my health, it's due to my age.

That may not be true for visual impairment, but I still think that question of how patients are interpreting these questions in relation to their health is a problem. When I administer the SIP, I stop at every single item and make patients think whether it is due to their health or not, because some limitations may be due to things that they don't consider health.

**DR. WARE:** I think an essential component of the research agenda that this group launches in its trials should be the systematic study of health and non-health attributions for limitations and distress and specific attributions to visual impairment as opposed to other health problems. This issue has to be systematically sorted out, and it's going to have a major impact on results.

**DR. BERGNER:** Dr. Patrick talked about responsiveness and a relationship to clinical measures. You were talking about the specific measures and you said that they tended to be more responsive and more related to the clinical measures.

But one also has to be very careful not to develop a measure or eliminate items from existing measures to such a degree that only items highly related to clinical measures remain. Because then one would have an instrument that says the same thing as clinical measures, and that would be superfluous. On the one hand they should be related but not so related as to be redundant.

Dr. Patrick also made the point that sometimes the longer generic measures can't be administered with very sick respondents. Experience with the SIP has been that it is possible. It depends on what you mean by very sick. Obviously, if a patient is cognitively impaired, none of these measures can be used. But they have been used with patients who are in nursing homes. I am currently using SIP in a study of very sick people who are hospitalized. When I say very sick, they have a mortality rate of 50 percent within 6 months. I know that the National Heart, Lung, and Blood Institute (NHLBI) has used it with oxygen-dependent chronic obstructive lung disease patients.

**DR. FRYBACK:** Back to your comment about a need for the quality table, or the growing popularity of the quality table. It seems that you can move generic items into a specific context or supplement them or add to or subtract from them and maintain most of the reliability and also the ordinal properties. But I think that the metric properties are quite fragile as you go from one population to another.

I would be concerned about the equal interval properties that are necessary for those sorts of quality computations, as you go from a generic base into quite specific populations. The interval properties may be quite different in the very, very sick population than in a less sick population. The step sizes and ranges are quite different.

**DR. PATRICK:** I think this is a question of how the metric is obtained. I don't think the way in which the metric is obtained affects the way in which it's analyzed with the specific population. One may have to do transformations because distribution is so highly skewed. But I don't think that the actual metric property of that scale has changed in its application.

**DR. FRYBACK:** In terms of proportions of respondents, one can do transformations in categories. But in terms of equal steps of impact on quality of life, which is what we're talking about with QALYs (quality-adjusted life-years), I'm not at all sure that one can. I don't see specific population testing. I see more generic testing. I don't see retesting.

Reliability tends to transfer quite well, I think, and I would be quite secure in the ordinal properties. But I think that the metric properties are much less secure, not in terms of distributions of responses but in terms of the equal step, the impact, the exact interpretation one needs to compute a QALY.

**DR. BERGNER:** The tables of QALYs assume that the values that are placed on health states are similar from population to population. The only piece of evidence that this is so comes from Balaban, and he didn't look at the metric properties. He looked at the values of the Quality of Well-Being (QWB) items in a population of arthritic patients and compared them with those in the general population and found out that the values or the weights that are placed on the items are the same. That doesn't deal with the metric problem, but one could use those data to take a look at the metric problem.

**DR. FRYBACK:** It's a concurrent problem that needs to be dealt with, I think, in the testing and validation. It shouldn't be forgotten.

**DR. PATRICK:** I agree that if one looked at the metric properties of all of those things used to compute QALYs, one would find they are very different as well.

**DR. FRYBACK:** Well, most of the literature on QALYs consists of just bare guesses.

**DR. PATRICK:** If we are using QALYs prospectively, then we're talking about estimations that could provide a means of planning. But a QALY table, in my opinion, certainly does not provide a basis to make decisions. For example, I am not going to decide to withhold hemodialysis on the basis of a QALY table. There are enormous dangers in setting priorities for different types of interventions. A QALY table is not going to decide which patients are going to get which treatments. That's the point I want to make. If we are going to start using QALYs for that, we are in very serious trouble.

**DR. WARE:** I think the point that was made yesterday by Gordon Legge is the major threat to measurement. That is, a complete shift in values may occur due to experience with a condition, and this shift essentially pulls the rug out from under standardized measurement. In other words, if I occupy a disease state that I've never occupied before and I relate that to the choices that I have in my life, I make a complete shift in my values in line with that experience. We saw that in the Evans' work in end-stage renal disease when standardized life satisfaction measures were used. Adjusted scores didn't differ for these four very ill groups compared to the general population. Given the choices that these people had, they had a major shift in their values.

Evans also made the point that people make adjustments. We adapt. All of these function measures are sliding scales; they measure the individual and the challenge that the individual's environment makes.

**DR. PATRICK:** This is true with every single study. We don't come with clean cohorts of patients that have just been diagnosed or are just at the beginning of the disease process. We try to do that in clinical trials. But every cohort study, and I consider clinical trials a special kind of cohort study, is a mix. The age or the time of onset is a very important variable.

**DR. FERRIS:** Our cohort studies, at least some of them, specifically take people who are not affected with a given disease and follow them as it develops.

**DR. PATRICK:** People adapt, perhaps, to dressing and eating and some of those activities, but they do not adapt to the difficulties of finding work or the problems of transport and access. Those types of things, which are definitely related to their condition, have some importance, and we need that broader focus.

**DR. RAND:** My primary concern is the variability of a lot of these measures considering the variability in the psychosocial milieu that any individual patient experiences. For example, as clinicians we might work in an emergency room where we can see two people who present the exact same problem. One is sitting there stoically and waiting his turn and the other is making such a ruckus that nobody can deal with anybody else.

With the subjective kind of measures and the self-reporting that we're dealing with here, that variability might have a significant impact on the outcome.

Also, I have had a clinical impression that some of this psychological overlay does impact on the patient's ability to be treated, first on responsiveness to instructions, and second, possibly on physiologic responsiveness to treatment. In a clinical trial, psychological factors may be closely related to outcome and influence the measures that you are talking about. Do you think some kind of measurement of those phenomena might be needed to provide a basis on which to interpret some of this other information?

**DR. PATRICK:** If we're talking about affective or psychological functioning, it's a component of many of these measures and may show variation. My experience with inflammatory bowel patients is that it does. On the physical categories of the SIP, we don't get a lot of variation. On the psychosocial categories, we get more. And on the specific measure, we get an enormous amount, because those patients that are affected by minor symptoms are very afraid. The concerns instrument shows a lot of variability. The answer is yes, you need to include those types of measures.

**DR. RAND:** Are they included? Is there a scale?

**DR. PATRICK:** We're not including trait measures or personality measures as part of health status outcomes.

**DR. McDOWELL:** That's been done in the pain area, though. An example is the behavior questionnaire by Pilowsky. There are several instruments that attempt to address the emotional aspect of the pain response in contrast to the more underlying physiological response, doing it by various scaling approaches or doing it by more psychological items which might identify a hypochondriacal response, for example.

**DR. PATRICK:** Some trials have included these things as effect-modifying variables in the trial itself.

**DR. WARE:** First of all, Larry Rand's point is very clear with respect to measures of psychological variables. But it would be a mistake to think that in measures of physical function and social and role behaviors, that there isn't any personality involved. There is a lot.

I think the defense against this problem, this threat to validity of interpretation, is a case-mix measurement. You have to do it on the clinical side, and you have to do it in testing functional status and well-being, even more so in the case of the latter, because of the sensitivity of these measures to a whole host of things.

So the first defense is thorough case-mix measurement. The final refuge, the salvation of science, is randomization. When you want to compare equivalent groups, to make sure that there isn't something about the individuals that brought a different outcome to the comparison as opposed to the intervention, you randomize.

**DR. RAND:** I think that in the randomized clinical trial that would certainly be a very important factor. But in other situations where you are looking at the successful outcome versus nonsuccessful outcome, and trying to look at their determinants, that would be much more problematic.

**DR. JAVITT:** You said something toward the end of your talk that excited me greatly, because I haven't heard it associated with a health status measure before.

Several years ago in a cataract study in the Third World, I made an empirical observation of people's perceived utility of having their sight restored. Although I knew nothing about scales of impact at that time, so I didn't have a way of reporting it, I had the sense that people's perception of the utility of regaining their sight was directly related to their economic productivity prior to losing their sight. That is, people who had been gainfully employed in one way or another perceived a greater benefit from regaining their sight than people who had not been gainfully employed prior to their disability.

I wonder whether there is a way of incorporating monetary measures associated with disability in terms of pre-disability income, income while disabled, and social services and other entitlement programs.

**DR. PATRICK:** Yes. And I think that they have been. We saw an interesting presentation yesterday from Peter Bernth-Petersen in which the new measures that he included to address policy were exactly such measures—return to work, disablement pension, and so on.

The other direction that has been taken, although I am not enamored with it myself, is the willingness-to-pay question. How much would you be willing to pay for the restoration of your sight? This puts it in clear monetary terms; it gives it a metric.

**DR. JAVITT:** People from different financial backgrounds have very different ideas of willingness to pay. People who know what a million dollars is are more likely—or less likely perhaps—to be willing to pay a million dollars.

# Incorporating Health Status Measures Into Clinical Trials

John E. Ware, Jr., Ph.D.

The focus of outcome assessment in clinical practice, health policy studies, and health services research is expanding beyond clinical measures. Although they have served the field well for a hundred years, clinical measures are incomplete. These measures reflect the values of clinicians, their diagnostic tools, and their highly refined assessments of the severity of very specific conditions.

They do not, however, reflect the values of people. What people value is their ability to function physically and mentally. What is important to them is their cognitive and affective states with respect to their health.

The term, quality of life, takes in more than health. In fact, according to one study, health is only one of about 12 attributes that compose the quality of an individual's life. As we broaden the concept of quality of life to include more than health, we are also broadening our focus to include more than the individual. We are no longer measuring only an individual's health status, but also that individual's social and community environment and even such issues as whether the country is at war, whether it is polluted, and so on.

This paper discusses differences between quality-of-life measures and offers criteria to consider in selecting instruments for use in clinical trials.

## Differences in Quality-of-Life Measures

Different measures focus on different health concepts and different aspects of those concepts. In addition, some measures assess behavioral dysfunction, while others depend on subjective perceptions of performance. These differences affect the findings of measurement.

Table 1 shows a selection of items that measure five generic health concepts. There is at least one perceptual and at least one behavioral measure for each of these concepts. Personally, given a lack of a better understanding of disease impact and treatment impact, I favor a measurement methodology that includes most of the items on this grid until we know more about what we do and what we don't do in each of these areas.

Some of the measures that we have talked about up until now, such as the Sickness Impact Profile (SIP), which is one of the most comprehensive measures you could think of, favor behavioral dysfunction. So if I were going to supplement the SIP with something else, it would not be with another measure of the behavioral dysfunction type. It would be something more in the well-being or the personal evaluation of health type of approach. Therefore, adding other items from this grid is an important thing to do if one is going to field a comprehensive health assessment.

## Criteria for Selecting Quality-of-Life Instruments

### *Comprehensiveness*

No one knows enough about the impact of a particular disease or treatment to choose among the health concepts most important to patients. To leave out any one of them would be a very risky thing to do.

While comprehensiveness is important in assessing quality of life, it may pose difficulties in practice. If respondent burden is a problem, one solution is to select a short form measure in areas that do not need to be measured in great depth. A short form measure can pick up a moderate or large, but not a small, effect. One advantage to

## Table 1
## Selected Items Measuring Generic Health Concepts

| Concepts | Definition | Abbreviated |
|---|---|---|
| PHYSICAL | | |
| Physical limitations | Limitations in performance of self-care, mobility, and physical activities. | Needs help with bathing, dressing.[1]<br>In bed, chair, couch, for most of the day.[2]<br>Do not walk at all.[3] |
| Physical abilities | Ability to perform everyday activities. | Able to walk uphill, upstairs.[4]<br>Able to participate in sports, strenuous activities.[5] |
| Days in bed | Confinement to bed due to health problems. | During past 30 days, number of days health keeps one in bed all day or most of day.[6] |
| Physical well-being | Personal evaluation of physical condition. | Rating of physical shape or condition.[7, 8] |
| MENTAL | | |
| Anxiety/depression | Feelings of anxiety, nervousness, tenseness, depression, moodiness, downheartedness. | Depressed or very unhappy.[9]<br>Bothered by nervousness or nerves.[7, 10] |
| Psychological well-being | Frequency and intensity of general positive affect. | Happy, pleased, satisfied with life.[7, 10]<br>Wake up expecting an interesting day.[11]<br>Feel cheerful,light-hearted.[12] |
| Behavioral/emotional control | Control of behavior, thoughts, and feelings during specified period. | Felt emotionally stable.[7, 10]<br>Lose control of behavior, thoughts, feelings.[12]<br>Laugh or cry suddenly.[3] |
| Cognitive functioning | Orientation to time and place, memory, attention span, and alertness. | Feel confused, forget a lot, make more mistakes than usual.[3] |
| SOCIAL | | |
| Interpersonal contacts | Frequency of visits with friends and relatives. | Number of friends visited.[13]<br>Going out less often to visit people.[3] |
| | Frequency of telephone contacts with close friends or relations during specified period of time. | How often on telephone with close friends/relatives, past month.[13] |
| Social resources | Quantity and quality of social ties, network. | Number of close friends, people to talk with.[13] |
| ROLE | | |
| Role functioning | Freedom from limitations in performance of usual role activities (e.g., work, housework, school) due to poor health. | Limited in kind or amount of major role activity.[6]<br>Working shorter hours.[3]<br>Health causes problem at work.[14]<br>Unable to work because of health.[5] |

## Table 1 [contd.]
## Selected Items Measuring Generic Health Concepts

| Concepts | Definition | Abbreviated |
|---|---|---|
| GENERAL HEALTH PERCEPTIONS | | |
| Current health | Self-rating of health at present. | In general, is health excellent, good, fair, or poor?[6]<br>Health is excellent.[15]<br>Energy, pep, vitality.[10]<br>Been feeling bad lately.[15] |
| Health outlook | Expectations regarding health in the future. | I expect to have a very healthy life.[16] |
| Pain | Ratings of the intensity, duration, and frequency of pain as well as limitations in usual activities due to pain. | During the three months, how much pain have you had?[1]<br>How much pain interfered with things?[17] |

References

1. Katz S, Ford AB, Moskowitz RW et al. Studies of illness in the aged. *Journal of the American Medical Association* 185:94-99, 1963.
2. Kaplan RM, Bush JW, Berry CC: Health status: Types of validity and the index of well-being. *Health Services Research* 11:478-507, 1976.
3. Bergner M, Bobbitt RA, Pollard WE et al. The sickness impact profile: Validation of a health status measure. *Medical Care* 14:57-67, 1976.
4. Hulka BS and Cassel JC. The AAFP-UNC study of the organization, utilization and assessment of primary medical care. *American Journal of Public Health* 63:494-501, 1973.
5. Stewart AL, Ware JE Jr, Brook RH. Advances in the measurement of functional status: Construction of aggregate indexes. *Medical Care* 19:473-488, 1981.
6. National Center for Health Statistics: Health. Washington, DC, US Department of Health and Human Services, 1981.
7. Dupuy HJ. The psychological section of the current Health and Nutrition Examination Survey. Proc Public Health Conf on Records and Statistics Meeting Jointly with the National Conference on Health Statistics. Washington, DC, National Conference on Health Statistics, 1972.
8. Chambers LW, MacDonald LA, Tugwell P et al. The McMaster Health Index Questionnaire as a measure of quality of life for patients with rheumatoid disease. *Journal of Rheumatology* 9:780-784, 1982.
9. Bradburn NM. The Structure of Psychological Well-Being. Chicago: Aldine Publishing Co, 1969.
10. Dupuy HJ. The psychological general well-being (PGWB) Index, in Wenger NK et al. (eds): Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies. New York, Le Jacq Publishing Co, 1984, pp 170-183.
11. Costello CG and Comrey AL. Scales for measuring depression and anxiety. *Journal of Psychology* 66:3030-3313, 1967.
12. Veit CT, Ware JE Jr. The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology 51:730-742, 1983.*
13. Donald CA and Ware JE. The measurement of social support, in Greenley JR (ed): Research in Community Mental Health. Greenwich, CT, JAI Press, Inc, 1984, pp 325-370.
14. Hunt SM, McEwen J, McKenna SP et al. The Nottingham Health Profile: Subjective health status and medical consultations. *Social Science and Medicine* 15A:221-229, 1981.
15. Davies AR, Ware JE Jr. Measuring Health Perceptions in the Health Insurance Experiment. Santa Monica, CA, RAND Corporation, R-2711-HHS, 1981.
16. Ware JE Jr. Scales for measuring general health perceptions. *Health Services Research* 11:396-415, 1976.
17. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin brief pain questionnaire to assess pain in cancer and other diseases. *Pain* 17:197-210, 1983.

picking up only moderate to large effects is that, if found, they are almost certain to be clinically and socially relevant.

### *Ability to Discriminate*

Measures also should discriminate among groups that are known to differ in health problems, such as different diagnostic groups, or the difference in severity among conditions within a diagnostic group.

### *Sensitivity to Change*

Another important issue in the evaluation of an intervention's impact is the sensitivity of the measure to change. Even for the most widely used and well understood measures, this is an issue that has yet to be fully explored. We do not know a lot about instruments' sensitivity to change in the context of particular trials.

### *Length of Scale*

A fourth attribute of a measure selected for a clinical trial is the length of the scale. This refers not to the number of items in the scale, but to range of the continuum along which items are place. To use an analogy, a rule of 10 feet is satisfactory only if everything to be measured is less than 10 feet. To learn about things that are longer than 10 feet, a greater range or continuum is needed.

This means that it is important to know where attributes to be measured fall on a continuum. If this is not known, then items should be at least roughly interspersed throughout the entire range, in order to have some detection capability. So length of the scale, or range of measurement, has an important impact on what we are able to learn from quality-of-life measures.

## Selecting Appropriate Measures

In suggesting that we use a minimum set of concepts or measures in a trial, regardless of the purpose of the trial, the five generic health concepts in Table 1 would be the ones that I would include in the set.

The reason is that regardless of which definition of health one uses, these are the things that come up again and again. They reflect physical, mental, and social functioning; some would argue that role functioning is the absolute bottom line from an economic or societal perspective. Also the person's own subjective evaluation of health is very important. In addition there are some very important concepts that are left out, such as pain. So this is an absolute minimum set.

I would argue further that we have erred in the development of this field on the side of using measures of specific concepts; that is, using specific measures of a particular concept that focus on one end or the other of the scale, to the point where we have missed a tremendous opportunity to create a data base that would answer such questions as "What are the differences in health between the general well population and people with a particular chronic condition?" or "What is the difference in health between someone who has diabetes, someone who has arthritis, and someone who has depression?"

As long as we use diabetes-specific depression measures for diabetes and arthritis-specific depression measures for arthritis, and depression-specific depression measures for depression patients, we are never going to be able to fill out this scale in terms of where these various groups are in relationship to one another.

Whereas there is a very strong case for specificity in measurement, there is also a very strong case for at least a limited standard set of measures that would be used regardless of age group, regardless of diagnosis, regardless of the severity within that diagnostic group, and regardless of treatment. And that is the kind of commonality in measurement that will help us move forward in this particular field. One of the limiting factors in executing that methodology is respondent burden.

## Respondent Burden

It takes over a hundred questionnaire items to measure these five generic health concepts using the better, lengthy, research versions of measurement scales. For most respondents, the time required is about 25 to 30 minutes; a personal interview takes a little longer. Many people, however, do take more time.

The Medical Outcomes Study, an observational study of variations in physician practice styles and patient outcomes in different systems of care, is evaluating two alternatives to lengthy research questionnaires. One alternative is to use the best psychometric methods to identify the most critical subset of items within each lengthy research instrument. The objective is to identify the subset that has the best predictable validity in relation to a variety of criteria.

Thus, rather than using a single item measure, of which we and others have been critical, we have compromised between the two extremes, using the long form measures as a criterion in the development of the most efficient, most cost-effective short form measure. This short scale approach reduced respondent burden by more than 80 percent in one case.

To carry this methodology further, we are evaluating ten poster charts developed by the Dartmouth Primary Care Cooperative Information Project that use words and pictures to depict health states. These were used in clinical practice for a couple of years in New England. There also are self-administered and telephone-administered narrative versions of these that do not require pictures. In fact, we are currently studying the value of the picture for different concepts.

The reliability of the measures in a random sample of adults in the United States, and among some 23,000 patients that we screened in the Medical Outcomes Study, is roughly comparable. But we are learning that reliability, although necessary, is far from sufficient. The real key to precision is the enumeration of a scale with a large number of categories. It is possible to devise a very reliable classification system that puts people into one, two, or three boxes. That level, of course, no matter high the reliability, is really not going to be adequate for detecting changes over time, because people can move around within a box to an unacceptable degree. The number of levels that are reliably enumerated is a much important criterion.

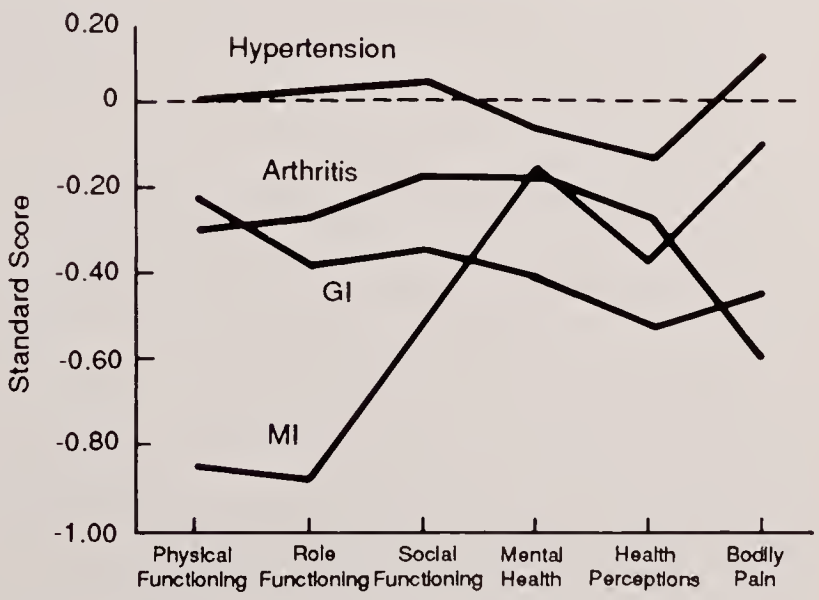
A fuller description of the Medical Outcomes Study is appended to this paper.

## A Health Profile

In the Medical Outcomes Study, in addition to comparing six general measures—physical functioning, role functioning, social functioning, mental health, health perceptions, and pain—we are looking across scales, to obtain a so-called health profile. For this particular comparison, we have omitted the general population group. Suffice it to say that it scores up above the zero level.

The mean score on each of these concepts for well patients, the 20,000 or so in doctors' offices that we sampled, is the dotted line across the top of Figure 1. So we call that zero. We then express any deviations from the mean for well patients as a proportion of a standard deviation on each of these particular scales. All six of these scales are scored favorably on a zero to 100 scale, and their standard deviations are in the 15 to 20 range. This is a way of standardizing comparisons for purposes of a health profile across very different metrics, using a metric that has a direct interpretation. By what proportion of a standard deviation unit does a particular diagnostic group differ on a particular health concept?

**Figure 1**
**Health Profiles**




Health profiles for patients with four conditions. Dotted line indicates patients with no chronic conditions; GI, gastrointestinal disorder; and MI, myocardial infarction.

Note the solid line for people with hypertension. Not surprisingly, they look very much like the well population. The only significant difference there is in general health perceptions. They have a three- to five-point decrement compared to well patients in their belief about their health. In other words, they are less likely to expect a healthy life, they are less likely to rate their health now as excellent, and they are a little more worried about their health. So cognitively, they have made an adjustment with respect to their health, but they do not behave or function any differently.

For arthritis, all of these differences are statistically significant from the comparison group. And the biggest decrement is in the area of pain. So arthritic patients, as expected, have a lot more pain and, as other studies have documented, there is a decrement in all of the areas of functioning, even in the well-being of the mental and perceptual areas.

For patients who have recently survived a myocardial infarction, there is a tremendous physical and role dysfunction. This is well-documented in the literature. The profile for type II diabetes, not shown on the graph, had a large decrement in health perceptions, no decrement in pain, and no decrement in mental health.

Let me take this opportunity to say that there is something basically wrong with the general mental health measures that we have been using to evaluate different conditions and treatments. We now strongly believe that one of the reasons we do not see differences in mental health is that when you ask people whether they are depressed or the frequency of symptoms of anxiety and so on, you do not get the same picture that you get if you ask them whether that despondency, despair, frustration, worry, or dread is specifically due to their health. We have compared the best mental health measures we know of with specific measures of distress due to health problems and we see a completely different picture, with health having a much greater impact. So again, this is an example of improving the validity of a measure with a more specific attribution to health.

## Conclusion

Quality-of-life measures should not replace clinical measures of severity. The two represent different concepts; they need to be measured and interpreted separately. But it is encouraging that the relationships between quality-of-life measures and clinical measures, at least on the average, are as strong as they are.

Although large differences exist at the mean, that essentially indicates that there is tremendous variation within even the most homogeneous diagnostic group. For people who have incurable chronic conditions, this is promising, because it shows that, although a disease or condition may set some upper limit on functioning in life, there is still a great deal that the individual can do. This leads to what may be the frontier in this field: new models of treatment. The analyses that I have outlined are based on the traditional medical model in which an intervention leads to some measure of improved function or well-being. There is much to be gained from this model, and these kinds of measures will document that over the next decade.

However, there are some patients who do very well in the face of their diseases and some that do very poorly. In the Medical Outcomes Study, these differences are not randomly distributed among the 526 physicians that we studied, nor are they randomly distributed among the systems of care that we sampled.

This finding has some important implications. We are trying to identify some of the variables that might explain why people have much better or worse outcomes than age or medical condition would suggest. The hypothesis here is that identifying these variables will lead to a new generation of interventions that deal more directly with the dysfunction and make it possible for medical providers to have even better outcomes with patients with relatively severe disease.

## Addendum: The Medical Outcomes Study

The Medical Outcomes Study is a study of variations in styles of practice and outcomes of care

for patients with chronic conditions. It is first and foremost a study of how to assess functional status and well-being routinely in a practice setting, e.g., what methods are most cost-effective, what intervals and what modes of administration should be used.

We also hope that the results will be sufficient to address some policy questions: Are patients with chronic conditions treated differently in different systems of care? Are there health implications of more intensive or less intensive resource utilization? What is the best way to measure case mix and are different groups selecting themselves into different systems?

We also are looking at specialty training to learn whether different specialty groups treat patients differently and if so, what the implications are for costs and outcomes of care.

The most important purpose of the Medical Outcomes Study, however, is to make methodological advances in the assessment of health.

The study was fielded in three cities, Boston, Chicago, and Los Angeles. We sampled three systems of care in each of those cities and 526 physicians from those systems. The physicians were in subspecialties associated with the chronic conditions that we are following.

We measured nearly 23,000 patients during an average of 9 days in each physician's office. We then picked a subset of patients who had one of a set of chronic conditions to follow over time. Independently of the physician and at periodic intervals, we are examining their care and measuring their functional status, using measures of well-being and clinical parameters. This is an older population and by design those eligible for Medicare are over-represented because of the obvious policy implications of that particular group.

One of the first things that we learned is that co-morbidity is the rule rather than the exception in doctors' offices around the country. More than 50 percent of the patients who had one of our chronic conditions had two or more. Co-morbidity is an important consideration and a good argument for a generic health assessment, because it is important to have a measure that is sensitive to a patient's entire profile of health problems.

So far, we have found that having two conditions is no worse than the sum of the two. Having diabetes and depression is essentially the sum of diabetes and depression decrements.

# Discussion

**DR. BERGNER:** As we discussed previously, people who have adjusted to an illness may not respond to certain items on a questionnaire because they no longer perceive a health problem. I've worked with hip replacement patients who no longer consider it a problem that they are contorting themselves in order to put their stockings on.

One of the reasons that the National Heart, Lung, and Blood Institute was very interested in functional status measures was the observation, which I think those of you in the eye field are making as well, that many patients with similar degrees of severity of disease, such as those with chronic lung disease, seem to live quite different lives. There is great variability in response to functional status measures, but that reflects a great variability in their response to their lives.

So it's not that the functional status measures are showing variability because they are unreliable. It's that people are individuals. One of the astonishments of the clinician is that a person who really is very sick may not act very sick.

So I don't think you want to get rid of the variation in whatever measurement you have. In fact, it would be better to retain it because it would give you information about what might be done or what areas might be the target of interventions—perhaps not purely medical interventions.

Donald Patrick was very concerned about the necessity, in asking questions about functioning, to emphasize strongly that the questions address functioning related to health. If health is not emphasized, people will not respond to certain items because they'll say, well, it's really related to my age.

But if you stress the effects of health too much, then what you do is force people to respond to something that they may no longer consider to be health-related. The adjustment is so well made that in fact the problem is no longer a health problem for them. So you have to weigh whether you want to push people into identifying everything as health-related or accept the adjustment they've made.

**DR. BAROFSKY:** In the paper just presented, do the data demonstrate no significant interaction in patients having two conditions?

**DR. WARE:** Yes. In an additive model, the estimate of the decrement due to depression and the estimate of the decrement due to diabetes, for people that have either of those conditions, is the same as those decrements in the person who has the two conditions simultaneously.

In the interactive model, having the two is worse than the sum of the two. That's the hypothesis, and from the point of view of clinical management, it makes a lot of sense because two diseases working together require a more complicated treatment, more often than not.

But out of 140 some tests, we had only three significant interactions. They were all in the hypothesized direction: Having the two conditions was worse than the sum of the two. But that wasn't enough to reject the additive model.

We're now looking at three conditions together, and, as you know, the data will get thinner. Someone else may do a study of 2,500 people with, say, depression and diabetes or hypertension and diabetes, and maybe they'll be able to see things that we were unable to see.

**DR. BAROFSKY:** But when you set up the two by two table with depression and diabetes, you did not find a significant interaction in that table?

**DR. WARE:** No. There are only three instances of interaction. We've got about as many significant interactions as we would expect by chance, having done enough comparisons.

So we don't see a lot of support for the interactive model at this point. But this analysis ignores severity. This is just a diagnostic group analysis. And we have some pretty noisy diagnostic groups here. We're now using clinical data to identify these people.

Given that diagnosis is a continuum, we have in our study a continuum of diabetic certainty. A diabetic is a diabetic when the doctor says it, the patient says it (there's tremendous disagreement there), and the patient injects insulin. It's surprising how many principal care physicians think their patients are diabetics when the patients don't and vice versa. So even diagnostic group membership in this study, even for the medical conditions, is in itself a continuum of definition.

**DR. BERGNER:** How did you define the patients in terms of their co-morbidities in this? Was it the patients' definition?

**DR. WARE:** This analysis uses both the doctor and the patient.

**DR. BERGNER:** So it has to be the intersection of those?

**DR. WARE:** We only required that in the conditions that we estimated. For all the other co-morbidities, we took any patient or doctor report of co-morbidity.

**DR. BERGNER:** And there's a whole list of co-morbidities, not simply the ones that you're studying?

**DR. WARE:** Right. We didn't try to estimate those because the study was not designed to do that, but we controlled for them statistically.

We're now looking at the clinical exam that we did to bring a third variable to the equation, because we're trying to understand how to do case-mix adjustment. And that's the reason for all the different methods of assessment.

# The Utility Approach to Measuring Health States

Michael F. Drummond, Ph.D.

The utility approach differs from most other approaches to measuring health-related quality of life in that it yields a single value for health states on an interval scale. On such a scale, death equals zero and complete well-being equals one. When combined with data on the years of life gained by health care interventions, the utility values for health states enable the calculation of the number of quality-adjusted life-years (QALYs) gained from therapy.

This generic measure of the output of health care programs has gained popularity with health economists because they can use it to make comparisons between interventions in terms of each intervention's incremental cost per QALY gained. Some argue that league tables of health care interventions, which show relative cost per QALY, can assist in decisions on health care priorities (Torrance and Zipursky, 1984; Williams, 1985). In the United Kingdom, a Department of Health and Social Security working party has applied the same logic (DHSS, 1986).

This paper provides a brief overview of the methods of assigning utility values and discusses the relevance of utility values to health care decision-making and to clinical trials.

## Methods of Obtaining Utility Values

There are three main methods of obtaining utility values, each with its strengths and weaknesses: These are the rating scale, the time tradeoff, and the standard gamble. Each method generates values by asking the respondent, who may be a patient, doctor, or member of the general public, to compare health states that are described to him or her.

Torrance (1987) has reviewed the various methods of utility measurement and argues that they are reliable, valid, and precise. And although utility values vary from individual to individual, group means are quite stable. This finding is important because the proponents of health utilities contend that utilities are most useful for making planning decisions, rather than individual treatment decisions. This issue is discussed at greater length below.

Utility measurements are undertaken usually in addition to, or instead of, other quality-of-life assessments. In one case, however, a general quality-of-life instrument, the Quality of Well-Being Scale, has been designed to generate health utility values (Kaplan et al., 1976).

## Health Utilities and Health Care Decision-Making

Drummond (1987a) has pointed out that two resource allocation processes occur in health care; the formal resource allocation process, as followed in health care planning; and the informal resource allocation process, as followed in the decisions clinicians make about the treatment of their patients.

Economists argue that utility assessments are most relevant to the formal health care planning process. This being the case, *whose* values are most relevant and how should these be obtained? For planning decisions one might argue that the values of the general public are the most relevant, because the community collectively provides much of the funding for health care in many countries. (In the United States, the Government provides about 45 percent of total health care funds.)

It might be argued, however, that the general public would have an interest in how patients value different health states. Experiencing a particular illness, after all, may be different from

responding to hypothetical scenarios as presented in the various methods of obtaining utility values.

In the case of clinical decisions for the individual patient, utility measurements can be viewed as an extension of the process of obtaining informed consent to treatment.

## Utility Measurement and Clinical Trials

The relevance of utility measurement in clinical trials varies, depending mainly on whether trials are intended to indicate the best treatment for the individual patient or the best treatment for the whole community of patients who are eligible for a particular therapy.

One of the conclusions from the research on health utilities is that different patients make different tradeoffs of risk and benefit. This led Paterson (1988) to argue that quality-of-life assessments were relevant in clinical trials because they assessed the net impact of efficacy and side effects. If one takes the view that trials are intended to indicate the best treatment for the individual patient, utility assessments within trials may be relevant.

On the other hand, would an individual contemplating a particular treatment have an interest in how those enrolled in the trial valued the trade off of risk and benefit? Or would this individual prefer to know the probabilities of events as assessed in the trial and then to make his or her own tradeoff?

Alternatively, if one takes the view that trials indicate the best treatment for the community of patients eligible for a therapy, then it may be relevant to undertake utility measurements during the clinical trial. For planning decisions, however, it might not be patients' valuations that are most relevant. Any relevant sample of the general public may suffice and utility measurements therefore be made quite separately from the trial.

One more issue is of interest. Is utility measurement within clinical trials ethical? It could be argued that the explicit nature of the tradeoffs explored during utility measurement might heighten the anxiety of patients randomly assigned to one or other therapy (Drummond, 1987b). It might affect enrollment in the trial or increase the number of withdrawals after randomization. However, this problem could be minimized if the utility measurements were performed only after therapy. In addition, it could be argued that if the informed consent procedure were thorough, utility assessment would induce no further anxiety.

## References

Department of Health and Social Security. Breast cancer screening. Report of a working group chaired by Professor Sir Patrick Forrest. London, HMSO, 1986.

Drummond, MF. Resource allocation decisions in health care: A role for quality-of-life assessments? *Journal of Chronic Diseases* 40(6): 605-616, 1987a.

Drummond, MF. Collaborative ocular melanoma study: Options for quality-of-life assessment. November (mimeo), 1987b.

Kaplan RM, Bush JW, Berry CC. Health status: Types and validity and the index of well-being. *Health Services Research* 11: 478-507, 1976.

Paterson, ML. Measuring quality of life in rheumatoid arthritis: The case of auranofin. *In:* G. Teeling Smith (ed), Measuring health: A practical guide. London, Wiley, 1988.

Torrance, GW. Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases* 40(6): 593-600, 1987.

Torrance GW and Zipursky A. Cost-effectiveness of antepartum prevention of Rh immunization. *Clinics in Perinatology* 11(2): 267- 281, 1984.

Williams, AH. Economics of coronary artery bypass grafting. *British Medical Journal* 291: 326-329, 1985.

# Discussion

**MR. BUXTON:** It has been found that kidney dialysis patients assigned a higher utility to the dialysis than people in the general population to whom the experience was described. These were not major differences, not completely different rankings.

We've done some very limited work with questions on the state of health after treatment for breast cancer. It's quite exploratory, really, and we were very cautious about actually approaching patients, until we got some experience with the methods.

We have now approached a small group of patients who were treated a year before. They are selected patients; access to them was through a clinician who gave us only patients who were well adjusted to their health state.

Their utility values, interestingly, were slightly higher but not hugely different from those of women to whom the states were described. We are going to be reporting those figures.

**DR. BAROFSKY:** Didn't the Torrance study include home-based dialysis versus hospital-based dialysis?

**DR. DRUMMOND:** I think that may be in the background paper. Their findings were that home-based dialysis was assigned a higher utility value than hospital-based dialysis. And I think continuous ambulatory peritoneal dialysis came out closer to the home dialysis than the hospital. That was in a subsequent study.

**DR. BERNTH-PETERSON:** Some years ago, economists talked a lot about cost efficiency and cost benefit. Is that the case now?

**DR. DRUMMOND:** This term "cost utility analysis," which refers to the type of analysis in which costs are expressed in money terms and benefits are expressed in these quality-adjusted life-years, is a relatively new term. I don't know who first coined it, but certainly we started using it in Canada a few years ago. And that branch of the field has had a bit of a resurgence.

But there is still quite a lot of interest in what you might call cost-effectiveness analysis which uses the traditional clinical outcome variables in the field concerned and looks at the costs. And that can be quite helpful within a given clinical field.

For example, when we did our cost-effectiveness study of treatments for cataract, we just looked at the kinds of clinical measures that we've heard about these two days, because then we weren't trying to compare investments in cataract treatment with anything else.

And so there is still a lot of cost-effectiveness analysis going on. I think cost-benefit analysis is less popular because people recognize that the easy-to-measure benefits, like loss of productivity, are less important. I mean, we've got 11 percent of the population unemployed in Britain. So who needs the health service to get people back to work? It doesn't have any real meaning. It may have meaning to the individuals themselves, certainly, getting their job back, rather than somebody else getting the job.

As for willingness to pay, a couple of interesting studies, mainly the auranofin study, found that arthritis sufferers were willing to pay an average of 22 percent of their household income for a cure for their arthritis, a finding which actually didn't vary so much by the wealth of the respondent. Apart from that, there's still not much interest in cost-benefit studies. I certainly don't spend much of my time on that. If I'm going in any direction from the cost utility approach, it's down to cost-effectiveness analysis rather than up to cost-benefit analysis. But that's just my preference.

*So taking quality-of-life measures one step further and trying to value them in dollars would be helpful for communicating the benefits of medical research programs.*

**MR. McMANUS:** The Office of the Director, NIH, is in the process of trying to develop examples of the benefits of procedures resulting from NIH research. For the purpose of the Director communicating with Congress, it would be quite helpful to be able to say that there were savings of X million dollars this year as a result of this or that scientific advance.

The use of QALYs is appealing, especially to an organization like the National Eye Institute, because maybe there are a lot of things that are missed in traditional economic analyses that could be included: quality-of-life measures and, indirectly, willingness to pay.

So taking quality-of-life measures one step further and trying to value them in dollars would be helpful for communicating the benefits of medical research programs.

**DR. DRUMMOND:** We are certainly not intending to throw out previous analyses with the bath water, because we say, if you look at the transcripts from Congressional hearings, it's quite powerful when the Director of the National Eye Institute or any other institute has said well, look, you've invested this money in this, and it has resulted in all these savings.

But we've got to educate the policymakers a little more. And there's a big problem, I think, in the United States with prevention programs; everybody is hooked onto this idea that prevention saves money. There have been one or two successes of this approach. The work that the Office of Technology Assessment did on pneumococcal vaccination and I think on influenza for the elderly shows dollar savings. However, I think that economists could do the field a disservice if we focus the policymakers too much on the dollar savings.

It's a matter of tactics, I suppose. But I'd like to do both. I'd like to try and get better measures of things that are easy to measure but also try and educate folks that there's more to this than just saving dollars.

**DR. BERGNER:** But you've got a problem. How are you going to put dollar values on quality-adjusted life-years? It's been because of the problem of assigning monetary values that concepts like quality-adjusted life-years are being developed, measurements that don't assign dollar values.

If you want to assign dollar values, you're back to classical cost-benefit analysis. Essentially, quality-adjusted life-year analysis is an attempt, to get away from the problems of cost-benefit analysis, which values everything in dollars. Or I guess I have to consider it a submethod of cost-benefit analysis, in the way I would say that cost-effectiveness analysis is also a submethod.

**MR. BUXTON:** I would agree with that. Certainly in the U.K., for whatever reasons, there are resistance and unhappiness in various camps about trying to put dollar or pound values on years of life, whether measured crudely or quality-adjusted. And the whole appeal and the whole point of the QALY approach is that it is a unit of measurement which avoids that dollar valuation issue, although it does get at the relative valuation, for use in the health care field, of different health states.

I agree, though, it doesn't address the question of how money should be spent. The choices, coming to that league table, have to be all health care interventions. You can't in practice put something like national defense as a choice into a QALY league table.

**DR. JAVITT:** No matter how much I were to say about nautical miles or minutes of arc, probably everybody in this room would measure the length of this table in either inches or centimeters.

And to an extent I think everybody in this room and probably everybody on this campus in one way or another is in the business of marketing improved health to people who are willing to purchase it. Most of us make the assumption that health care in one way or another generates improved health.

We can talk about the utility of improved health to the public or the marginal utility to the patient or the physician, but the people that we're really marketing improved health to, the people who purchase health, are either legislators or sometimes administrators or employers. I wonder whether any research has ever been done asking legislators, asking corporate CEOs, asking people who have to ultimately make the purchasing decisions about health care, what units of measurement they will accept, what units of measurement they are willing to believe in making those health decisions, and what units of measurement they consider to be intangible.

**DR. PATRICK:** I actually have an answer for that, empirically, in that I compared three different methods for obtaining utilities comparing the answers of consumers and decisionmakers in New York State in the late 1960s. And one of the methods asked them to trade off numbers of people. How many people would you keep alive for a period of time in a certain state of health compared to a standard of 100 in perfect health?

The consumers found that extremely difficult to do and the method failed. The decisionmakers, much to my amazement, were able to do it. Some people said I would trade 255,000 people with a minor cold for 100 people in perfect health. I could tell the public health decisionmakers from the ease with which they made tradeoffs, and they were used to thinking that way.

**DR. LOHR:** One other point: There is clearly a movement in the whole field of quality assurance and quality assessment to incorporate the concept of patient utilities as a valuable adjunct, if not an absolutely necessary element, in patient and physician decision-making about clinical options. In this field, the patient utilities won't be handled in the same way that they would be for other sorts of uses.

The concept, furthermore, is now beginning to be built into the conceptual framework of quality review and quality assurance from a bioethical point of view.

I did want to underscore the point made about examining the ethical implications of this. In the quality assurance field, you can make a fairly strong argument about the need for revised notions of beneficence in order to accommodate notions of patient autonomy. This gets you to an argument that says you must take patient utilities for different health states into account in the actual decision-making before somebody gets one therapy rather than another.

Dennis Fryback and the people who were at the Institute of Medicine's Conference on Advances in Health Status Assessment last July in Menlo Park will recall the examples that were given of the choices for people who suffer from benign prostatic hypertrophy and whether to go with medical watchful waiting or in fact to have a prostatectomy. I want to underscore the idea that utilities and the whole ethical domain is coming much more to the fore in a completely different field.

**DR. DRUMMOND:** Well, I agree with a lot of what's been said the last few minutes. It's clear, isn't it, that the more we get into this question of valuation, the harder it gets. And I think the tradeoff is whether it's better to get a precise measure of something that's easy to measure or whether it's better to try and get less precise measures of things that are difficult to measure but may reflect the right concept.

# Major Themes and Conclusions: A Strategy for Measurement of Quality of Life In National Eye Institute Trials

Michael F. Drummond, Ph.D.
and
Frederick F. Ferris, III, M.D.

Above all, this workshop has shown that the measurement of health-related quality of life is a complex task that needs to be approached with due caution. Certainly it would be inappropriate to undertake quality-of-life assessments in clinical trials without an adequate appraisal of the suitability of existing instruments or the thorough validation of new instruments. The purpose of this concluding chapter is to outline a strategy for measurement of quality of life in National Eye Institute-supported clinical trials, drawing on the ideas expressed during the workshop.

First, quality-of-life measures should not be viewed as a replacement for traditional clinical measures of treatment success, such as visual acuity and visual field. These items are easy to measure and relate to the main objective of eye care, namely to improve vision. Nevertheless, measures of quality of life and measures of visual function, as discussed by Legge (1988), could be included in selected trials alongside the traditional measures. This would enable the convergent validity of measures to be explored; namely, if a person's visual acuity is improved, does this also lead to improvements in functioning and quality of life? In situations where agreement between measures is demonstrated, future measurements can be restricted to those which are inexpensive and reproducible, such as visual acuity.

Secondly, not all trials will require a quality-of-life component. As indicated above, it may be possible to extrapolate from simple measurements of visual acuity. In addition, the relevance of quality-of-life measurements will depend upon the purpose of the trial. Certainly, quality-of-life measurements are likely to be necessary if comparisons are to be made between eye health care interventions and those for other diseases. This is particularly important if eye care is in competition for limited resources with other programs.

Thirdly, in situations where quality-of-life measurement is indicated, it would be unwise to begin by developing a completely new scale. The development and validation of instruments can take at least 18 months; and it may be that one or more of the existing general scales are relevant. For example, it was pointed out at the workshop that some seemingly disease-specific scales such as the Arthritis Impact Measurement Scale (AIMS) are really general scales that obtain their specificity by the addition of such statements such as "due to my arthritis" to the quality-of-life questions. A conclusion from the application of these scales is that the impact of disease on the patients is quite wide-ranging. One should not assume that the impact of eye disease on quality of life is limited to those items obviously related to vision.

Fourthly, the most promising approach would be to adapt or to extend one of the existing general scales, such as the Sickness Impact Profile (SIP). This would enable specific aspects of visual impairments to be explored, but still enable the quality of life of eye care patients to be compared with that of individuals suffering from other diseases. The selection of a particular general scale would depend on the relevance of items included, its ease of administration, and its generalizability.

Finally, in order to reach agreement on the relative value for money from investments in eye health care, utility measurements should be made for certain key health states relating to degrees of visual impairment. In combination with data on life expectancy and cost, these measurements will enable the cost per quality-adjusted life-year (QALY) of eye care interventions to be assessed and compared with those from other branches of medicine (Drummond, 1988). However, utility measurements do not necessarily have to be made on patients in a clinical trial. They could just as well be made on other individuals experiencing degrees of visual impairment or members of the general public presented with descriptions of what it would be like to be in a given health state. The descriptive quality-of-life data collected through the use of general quality-of-life scales in clinical trials will be useful in devising the health state descriptions.

In conclusion, this chapter has indicated the main elements of a strategy for measurements of quality of life in National Eye Institute studies. Although the inclusion of quality-of-life measures will greatly increase the relevance of clinical trial results to potential patients and the general public, it is important that they be introduced in a way that does not greatly increase trial complexity and cost. The goodwill of trial participants and clinical investigators needs to be maintained if more effective eye care interventions are to be developed and tested in the future.

## References

Drummond, MF. Economic aspects of cataract. *Ophthalmology* 95(8): 1147-1153, 1988.

Legge, G. (1988). This monograph.

# Appendices

# Appendix A.
# Measuring and Valuing Health States Relating to Visual Impairment: A Review Of Literature, Concepts, and Methods

B.A. Ferguson*
M.J. Buxton*
M. F. Drummond+

HERG Research Report No. 3 — January 1988

Health Economics Research Group
Brunel University
Uxbridge, Middlesex

* Health Economics Research Group, Brunel University
+ Health Services Management Centre, University of Birmingham

## Contents

**71** Introduction

**71** Clinical Measures of the Benefits of Treatment for Eye Disease

**76** The Measurement of Aspects of Functioning and Quality of Life in Existing Clinical Studies

**77** The Use of General Instruments to Measure Quality of Life of Visually Impaired People

**80** Measuring Utility of States of Visual Impairment

**85** Particular Problems in Using Existing Measurement Techniques with Visually Impaired People

**88** Conclusions and Recommendations

**91** References

***Acknowledgements***

This paper is based on a report prepared for the National Eye Institute (NEI), National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, U.S.A. The authors are grateful to the NEI for financial support under contract number 263STH15486 and the constructive comments made by its officers. We are also grateful for comments by Barbara Hawkins on an earlier draft. However, all remaining errors are the sole responsibility of the authors.

## Introduction

There is an increasing interest in assessing the success of medical interventions in terms of their impact on quality of life. This grows from a recognition that patients themselves are not interested in improvements in a biomedical indicator but in how treatment affects their mobility, ability for self-care, and socio-emotional status.

A large range of measurement instruments now exists and these have been used both to assess the health status of populations and to examine the impact of particular interventions such as heart transplantation (Buxton and coworkers, 1985; Hunt and coworkers, 1986). Similarly, there is a growing body of research focusing on the values placed by patients and members of the general population on particular health states (Torrance, 1986).

The clinical literature on treatment for eye disease has continued to focus almost exclusively on measures of visual impairment, with only limited consideration of broader measures of health status. Frequently, the assessment of treatment effects is limited to visual acuity and visual field.

In a recent paper commissioned by the National Eye Institute, Drummond (1986) discussed a particular form of health status measure, known as "utility." He pointed out that there was a growing tendency to rank health care interventions in terms of their cost per quality-adjusted life-year (QALY) gained and made some preliminary estimates of the QALYs gained from cataract extraction.

To date there has been no empirical study of the utility values for states of visual impairment that could feed into such considerations of cost per QALY. In the future, research-funding agencies such as the NEI will have to make decisions on whether they will support the inclusion of quality-of-life-measures in clinical trials and assessment of utility values for various states of visual impairment. On the one hand, such measures will be useful in that they will facilitate comparisons of interventions in the eye care field with those in other branches of medicine. (This may illuminate the debate on priorities for health care and medical research.) On the other hand, little is known about the difficulties of making these measurements and it would therefore be unwise to make a major investment without some prior investigation.

***This paper:***

- Reviews some of the clinical literature on treatment for visual impairment, focusing initially on the way that clinical benefits of treatment for eye disease are typically measured (a selective annotated bibliography is available as an Appendix B to this publication).
- Discusses the limited consideration of broader aspects of benefit within the literature and the evidence for the relationship between quality of life/functional limitation and the normal clinical measures.
- Focuses on the general instruments available that might be used to measure the quality of life of visually impaired people and discusses the relevance and applicability of utility measurement.
- Briefly considers whether the use of measurement and valuation instruments with visually impaired subjects involves any major problems.
- Concludes with a series of recommendations for appropriate steps to begin to rectify the current paucity of research in this field.

## Clinical Measures of the Benefits Of Treatment for Eye Disease

In reviewing the clinical literature on eye disease and its treatment, visual acuity emerges as the prime clinical measure of visual performance, and thus changes in visual acuity (or differences in visual acuity between groups) are the major indicators of patient benefit. For example, the British Multicentre Photocoagulation Trial (1975 and 1977), concerned with treatment for diabetic maculopathy and proliferative diabetic retinopathy, presents its results in terms of visual acuity, arguing that:

> "to the patient....retention of visual acuity is the most important index of successful treatment, and....this functional measure provides the most

> valid basis for therapeutic or prophylactic photocoagulation in diabetic retinopathy." [1975, p. 1113]

Similarly in the U.S., the preliminary report of the Diabetic Retinopathy Study Research Group (1976) measured treatment effects solely in terms of visual acuity and visual field. The 5-year followup (Diabetic Retinopathy Study Research Group, 1984) used visual acuity alone as the "principal response for gauging the efficacy of the treatment" (p. 374). More recently, the Macular Photocoagulation Study Group (1986) again used visual acuity as the primary outcome variable.

Visual acuity has the advantage of being relatively easy to measure without sophisticated equipment. It can also be measured by unskilled personnel, although if best-corrected visual acuity is to be measured, obtaining the correct refraction may require more skill. Also, measurement involves no inconvenience, discomfort, or distress to the patient. It can be, and frequently is, converted to a simple point scale (see Table 1).

**Table 1**
**Coding for Visual Acuity**

| Code | Snellen |
|---|---|
| 1 | 6/5 or 6/6 |
| 2 | 6/9 |
| 3 | 6/12 |
| 4 | 6/18 |
| 5 | 6/24 |
| 6 | 6/36 |
| 7 | 6/60 |
| 8 | 3/60 and Count Fingers |
| 9 | Hand Movement and Perception of Light |
| 10 | No Perception of Light |

Source: British Multicentre Photocoagulation Trial, 1975, Table 1, p. 1116.

Such a scale can be easily subjected to statistical analysis. Strictly, the resultant scale is simply ordinal, that is, a ranking that gives no true indication of the magnitude of the intervals between points. Mean scores must be interpreted with extreme caution, and nonparametric statistical tests should be used. A one-point change in score at opposite ends of the scale may be of quite different practical significance, or utility, to the subjects concerned.

The issue of scaling visual acuity measurements has been discussed in detail by Westheimer (1979). It is possible to arrive at a single value for the visual acuity of an eye under a given set of conditions. This single value is a threshold, known as the "minimum angle of resolution" (MAR), measured in minutes of arc (where "normal" MAR is taken to be 1 minute of arc). Westheimer examines four methods of scaling visual acuity, recommending the logarithmic scale of MAR as a good approximation to an equal-discriminability scale for visual acuity. With the LogMAR values, an interval scale is generated, whereby the steps from (for example) 6/6 to 6/12 and from 6/12 to 6/24 are equal. The MAR values themselves are simply the reciprocal of the decimal Snellen fraction; for example, the MAR corresponding to a Snellen visual acuity of 6/12 is equal to 12/6 = 2 minutes of arc. An example of a scale illustrating these various elationships is shown in Table 2. Note that the decimal scale is nonlinear, whereas the LogMAR scale is linear.

Recent clinical trials (e.g., the Early Treatment Diabetic Retinopathy Study) have utilized new visual acuity charts that lend themselves well to the above log-linear scale. Ferris and colleagues (1982) pointed out three important deficiencies in the conventional Snellen-type charts for measuring visual acuity. First, many of the standard charts have a different number of letters on each line, hence allowing one mistake per line has a different meaning at different levels of visual acuity. Second, most charts do not have a regular progression in letter size from the easiest to the most difficult lines, making it difficult to interpret a change in visual acuity measured as "number of lines gained or lost." Third, Snellen letters vary widely in difficulty.

## Table 2
## MAR Measurement of Visual Acuity

| Snellen Visual Acuity | Decimal Fraction | MAR | LogMAR |
|---|---|---|---|
| 6/6 | 1 | 1 | 0 |
| 6/12 | 0.5 | 2 | 0.3 |
| 6/24 | 0.25 | 4 | 0.6 |
| 6/48 | 0.125 | 8 | 0.9 |
| 6/96 | 0.0625 | 16 | 1.2 |
| 6/192 | 0.03125 | 32 | 1.5 |

Ferris and coworkers designed new visual acuity charts to take account of these limitations. The new charts would have the same number of letters on each line (5), a regular progression of letter size from line to line, and use letters of approximately equal difficulty, ensuring that each line has approximately the same overall difficulty. This last objective was achieved by using the 10 Sloan letters—S, O, C, D, K, V, R, H, N, Z and assigning a degree of difficulty to each, according to how often that letter is read correctly at the visual acuity threshold. A line-difficulty score was then calculated for each of the 252 combinations of 5 letters that can be created from the 10 Sloan letters, by summing the five letter-difficulty scores. Finally, 28 lines of intermediate difficulty were selected for inclusion in two separate charts (14 in each), one chart to be used for each eye.

A LogMAR scale can then be constructed (see Table 3) such that a move from any one line to a consecutive line is represented by a 0.1 LogMAR unit difference. One example of this log-linear scale is the relative accuracy that it provides if incomplete lines are read correctly by the subject. Since there is a 0.1 LogMAR unit difference between lines on the charts, and since each line has five letters, each letter read correctly on the chart can be assigned a score of 0.02 LogMAR units. Indeed, "By scoring 0.02 for each letter read correctly on the entire chart and adding these scores, a visual acuity score can be created that is a single number and is a reasonable estimate of the LogMAR score for that eye at that test" (p. 96). Clearly, a further advantage is that changes in visual acuity scores over time can easily be subjected to statistical analysis.

## Table 3
## Log-Linear Scale of Visual Acuity

| Snellen Visual Acuity | LogMAR |
|---|---|
| 6/60 | 1 |
| 6/48 | 0.9 |
| 6/38 | 0.8 |
| 6/30 | 0.7 |
| 6/24 | 0.6 |
| 6/20 | 0.5 |
| 6/15 | 0.4 |
| 6/12 | 0.3 |
| 6/10 | 0.2 |
| 6/7.5 | 0.1 |
| 6/6 | 0 |
| 6/5 | – 0.1 |
| 6/3.75 | – 0.2 |
| 6/3 | – 0.3 |

Source: Ferris et al., 1982, Table 3, p. 95.

The issue of scaling becomes more complex if other aspects of visual performance are considered. Colenbrander (1977) proposes a seven-point scale of visual impairment incorporating both visual acuity and visual field (see Table 4). The Early Treatment Diabetic Retinopathy Study Research Group (ETDRSRG) (1985), while again using visual acuity as "the primary means of gauging treatment effects," incorporated other formal tests of visual function and also recorded patients' impressions of change in their visual function.

There has been a number of suggestions within the ophthalmological literature that visual acuity (even when supplemented by other measures of visual functioning) may be inadequate as an indicator of degree of visual impairment. Genensky (1976) argued that visual acuity measurement does not tell us very much about what a patient can expect to accomplish with his residual vision. For example, it does not necessarily say anything about the quality of the image perceived by the patient. Colenbrander (1977) distinguishes visual disability (relating to the ability to perform tasks visually) from visual impairment (relating to visual functioning) and notes that while they are related there is not a "simple one-to-one relationship" (p.OP335). He points out that "unfortunately, no accurate measuring scales are available" for visual disability "like those we have for visual acuity." He goes on to propose a simple classification of levels of visual disability (see Table 5).

Comparison of Tables 4 and 5 raises a further issue. Namely, measurements of visual acuity are usually reported on one eye, usually the worse eye. However, the ability to function is affected by the sight in both eyes. This is important to note in assessing the improvements in quality of life from interventions in the eye care field, since the quality-of-life gains from improving the visual acuity in one eye will inevi-

**Table 4**

**Levels of Visual Impairment**

| | | |
|---|---|---|
| No impairment | Visual acuity ≥ 20/25 and visual field ≥ 120° | |
| Slight impairment | Visual acuity < 20/25 or visual field < 120° | |
| Moderate impairment | *Visual acuity < 20/60 or visual field < 60° | Eligible for educational assistance (USA) |
| Severe impairment | *Visual acuity < 20/160 or visual field < 20° | Legally blind in USA eligible for tax deduction and other benefits |
| Profound impairment | *Visual acuity < counting fingers at 3m or visual field ≤ 10° | Blindness as defined by WHO and several European countries |
| Near total impairment | *Visual acuity < counting fingers at Im or visual field ≤ 5° | Blindness as defined in several developing countries |
| Total impairment | *No light perception | Blindness as defined in dictionary |

*Criteria are based on recommendations of the WHO Expert Committee on Prevention of Blindness (1972) and the US definition of legal blindness. Other values are extrapolations suggested by the Committee on Terminology, International Council of Ophthalmology.

Source: Colenbrander, 1977, Table 2, p. OP334.

tably depend on the visual acuity in the other eye, and on whether (if the patient is suffering from bilateral disease) this is likely to deteriorate.

Bernth-Petersen (1981) similarly suggests that it is difficult to equate visual acuity measurements with the patient's ability or inability to function visually and that a standardized measure of visual functioning is needed.

Alexander and coworkers (1987) examined the interrelationship between visual acuity, "peak contrast sensitivity," and the ability to perform certain visual tasks (five in total) in a group of low-visual-acuity patients suffering from age-related macular degeneration. Their data showed that patients had difficulty performing the simplest of tasks, and that these difficulties increased with deteriorations in visual acuity or contrast sensitivity. However, since visual acuity was correlated with contrast sensitivity, multivariate techniques were required to separate the confounding effects. Results were less clear when this was carried out; for three of the tasks involved, treated eyes had lower scores even after adjusting for contrast sensitivity. This perhaps emphasizes the need for a range of measures of visual function when assessing individual patients or evaluating particular treatments (e.g., laser photocoagulation). Indeed, the authors point out that "contrast sensitivity and other measures of visual function, as well as degree of visual acuity loss, should be considered in order to characterize more fully the visual capabilities of the eye."

These reservations about the use of visual acuity in isolation emphasize the problems of interpreting visual acuity scales derived from chart testing. With many of the conventional charts, there is no basis for assuming that movements between any two consecutive lines carry equal weight. At least with the "new" charts discussed above (Ferris, 1982), one can have more confidence that movements between lines can be regarded as representing equal changes in the ability to see. However, with any of the resulting scales—whether nonlinear decimal or log-linear scales—it is not clear that movements between any two adjacent points on the scale represent equally valued (from the patient's perspective) changes in the ability to see or, more generally, to function.

Finally, it should be noted that patients suffering from visual impairment frequently suffer other health problems—for example, elderly diabetic patients with additional renal and foot complications. This will also affect the quality-of-life gains

**Table 5**
**Levels of Visual Disability**

| | |
|---|---|
| No disability | |
| Slight visual disability | Can perform visual tasks without special aids (may require reduced reading distance, reading adds up to 4 D) |
| Moderate visual disability | Can reach near-normal performance with special aids (magnifiers, reading adds over 4 D) |
| Severe visual disability | Can perform visual tasks with aids but at a reduced level (reading speed, reading endurance reduced) |
| Profound visual disability | Cannot perform most detailed visual tasks (reading), experiences difficulty with gross visual tasks (mobility), has increased reliance on other senses |
| Near-total visual disability | Vision unreliable; relies mainly on other senses |
| Total visual disability | No vision; relies on other senses entirely |

Source: Colenbrander, 1977, Table 3, p. OP335.

from improvements in visual acuity and reinforces the need to take a broader view of the interrelationship between visual functioning and general aspects of quality of life in daily living.

## The Measurement of Aspects Of Functioning and Quality of Life In Existing Clinical Studies

In considering the broader impact of visual impairment, a very wide range of aspects of quality of life would seem a priori to be potentially relevant. These may encompass physical mobility, ability for self-care, and sociopsychological dimensions. A small number of these have been addressed, if only partially, in the clinical literature.

Psychosocial impact has been considered in a number of studies. Jastrzembska (1973) reviewed the few studies then published. An exploratory study by Calek (1973) of just 11 severely visually impaired subjects identified a number of aspects of emotional strain experienced by the subjects. Oehler-Giarratana (1980) reported a detailed study of 33 adult diabetic retinopathy patients. The study, using a semi-structured interview and two standard psychological instruments (Zung's Self-Rating Depression Scale and the Minnesota Multiphasic Personality Depression Scale), found significant evidence of severe depression associated with restrictions to physical mobility or role functioning.

A more recent review of the literature (Wulsin and coworkers, 1987) identifies a number of specific psychosocial problems, including psychiatric symptoms, among diabetic retinopathy patients. It makes the potentially very important suggestion that partial visual impairment appears to cause as much psychosocial disruption as severe blindness. The reasons for this are not clear. One may speculate that the partially sighted person is in a difficult marginal social position, since he/she functions neither as a sighted nor as a blind person. It may also be the case that partially sighted people deny to themselves or others the extent of their disability, thereby creating conflicts between expectations and performance. This view appears consistent with the findings of Ash and colleagues (1978), who found that subjects reported greatest psychosocial disturbance at the onset of visual loss rather than at the time of blindness registration. But in both these papers the tentative conclusions have to be tempered by the observations that there is a paucity of existing research and that further research is needed.

Similarly, the relationship between clinical measures of visual performance and functional or occupational ability has been little studied. Burack (1965), in presenting an interdisciplinary classification for the aged and applying it to residents of a home for aged blind people, does not "consider blindness as a major handicap influencing the ability to function." However, this rather surprising statement has to be tempered by the fact that the case illustrations all refer to elderly people with considerable medical and functional problems. In other words, the additional handicap of blindness may be relatively unimportant given their existing health problems. For example, one patient (aged 86) was "bedridden and incontinent with complete absence of independent activity." The handicap of blindness to otherwise active members of the working population must be considered in a totally different context.

It is also crucial to note that Bernth-Petersen (1982), in assessing the outcome of cataract surgery, reported the following. Twelve of the retired patients in his study stated that "poor vision would have forced them to give up living in their own homes and instead live in geriatric nursing homes." In ten of these cases, objective measures of visual acuity, visual functioning, and degree of support from family/friends supported the patients' subjective statement.

The inconsistency of the relationship between measures of visual acuity and the extent of handicap experienced is one of the main conclusions from a survey studying 193 visually disabled people in the community (Cullinan, 1977). It includes an analysis relating visual acuity to mobility of the subjects; within the limits of a small sample, it supports the hypothesis of no relationship. However, this analysis uses a very insensitive classification of mobility.

The work of Bernth-Petersen (1981) suggests that, in fact, there is a positive correlation between visual acuity and visual functioning but that the relationship is not a strong one. Unfor-

tunately, he provides no statistical measure of correlation.

Given the variety of handicaps suffered by some of the elderly visually impaired people who have been included in these relatively small studies (e.g., Burack, 1965), it is hardly surprising that a simple relationship has not been identified. The same may apply for other groups of visually impaired individuals, such as people with diabetes. The latter group frequently suffers from a variety of diabetes-related complications, such as neuropathy and renal impairment.

Studies considering broader aspects of the impact of eye disease have generally used nonstandard instruments or questionnaires. Thus the survey by Cullinan used a number of specifically chosen questions, which focused on specific vision problems and sight-related difficulties.

Genensky and coworkers (1979) devised a questionnaire codebook and mobility rating form which they hoped would serve as a guide for designers of new questionnaires for partially sighted populations. While it provides some important methodological tips, the questionnaire is aimed exclusively at the partially sighted and therefore cannot provide wider points of comparison. Interestingly, Haase and Bryant (1973) developed a distance scale for subjective measurement of perceived functional vision loss. It consists of five questions with Guttman scaling characteristics. It offers an interesting alternative to measurement of visual functioning by a clinician or other person but still focuses on the degree of vision loss rather than the impact of vision loss.

Bernth-Petersen (1981), recognizing the difficulty in equating visual acuity measurements with the patient's ability or inability to function, has devised a self-assessment index for measuring visual functioning. He defines visual functioning as the performance of everyday tasks that are vision-dependent, such as coping with work, housework, self-care, and mobility. The instrument has been used for the assessment of cataract patients, and consists of 11 questions relating to direct visual limitations, mobility limitations due to visual loss, and social role limitations due to visual loss (see Table 6). The scoring system was calculated to obtain the best discrimination among three distinct groups and, in fact, uses only questions 1, 2, 3, and 7. Hence the scale does not take into account the importance of the items to the patients, but merely the ability to discriminate among groups.

Thus, a number of studies have attempted to look at particular aspects of the impact of visual impairment but have tended to do so in very partial ways or have used specific, narrowly focused instruments, applicable only to visually impaired people. Some of the apparent inconsistencies in the conclusions that might be drawn from this literature reflect the complexity of the interrelationship between visual performance, visual impairment, and the impact these have on a patient's life style and quality of life, given that sight impairment is just one of many determinants of a patient's general health and social well-being.

## The Use of General Instruments To Measure Quality of Life Of Visually Impaired People

In an attempt to measure broader concepts of health and well-being, a number of so-called "general" or "global" instruments have been developed. These take into account as far as possible all aspects of an individual's quality of life. In the context of visually impaired people, this approach would look further than merely the immediate impact on sight of a particular disorder. It would attempt to identify—and also quantify—the impact on wider aspects of daily living such as self-care, mobility, and dependency. More sophisticated instruments, in addition, attempt to identify the effect on an individual's psychosocial and emotional state; for example, to what extent the disability affects social contact with others, or how the individual copes emotionally with having to lower aspirations and change lifestyle. These instruments are general in that they have not been devised for specific patient groups, diseases, or health problems, but can be applied generally to so-called healthy populations and to a wide variety of patient groups.

The recent study on cataract surgery with

## Table 6
## Index for Measuring Visual Functioning Levels

| Item Content |
|---|
| DIRECT VISUAL LIMITATIONS |
| 1. Reading capacities. What are you able to read in the newspaper? (nothing, large printing types, small printing types) |
| 2. How is your distance vision? (good, moderate, poor) |
| 3. Are you able to watch television? |
| MOBILITY LIMITATIONS DUE TO VISUAL LOSS |
| 4. Does your vision prevent you from automobile or bicycle driving? |
| 5. Is your vision sufficient for indoor orientation? |
| 6. Is your vision sufficient for outdoor orientation? |
| SOCIAL ROLE LIMITATIONS DUE TO VISUAL LOSS |
| 7. Does your vision limit the kind or amount of work or housework you can do? (major activities) |
| 8. Does your vision limit the kind or amount of other activities (e.g., shopping, gardening, hobbies, clubs) you can do? (other activities) |
| 9. Do you need help with eating, dressing, bathing because of your vision? (self-care activities) |
| 10. Do you need any help from the community to get along in daily life because of your vision? |
| 11. Do you need any help from your family to get along in daily life because of your vision? |

Source: Bernth-Petersen, 1981, Table 1, p. 199.

intraocular lens implantation (Applegate and coworkers, 1987) is unusual in that it recorded measures of visual acuity, and it also used a general measure—the Functional Assessment Inventory (Pfeiffer and coworkers, 1981). This is a multidimensional instrument, incorporating questions on activities of daily living (ADL), physical health, economic resources, emotional health, and social support. It was constructed for general use, not specifically with visually impaired people. Applegate and colleagues found that the surgery produced a positive improvement in some of these dimensions (for example, ADL), but these improvements were modest compared to the changes in visual acuity.

It would appear that the ADL scale, which has been extensively used (for example, in studying elderly patients), was not sufficiently sensitive in this case. ADL scales—of which there are many variants—typically assess individual performance in terms of a number of main areas. The version of Katz and Akpom (1976), probably the most widely used of these scales, involves classifying the patient as independent or dependent in bathing, dressing, going to the toilet, transfer (in and out of bed or chair), continence, and feeding. Because it has the features of a Guttman scale, with a normal hierarchy of severity, it produces a seven-point scale for "scale types" which conform to the standard Guttman characteristics (see

## Table 7
## Index of Independence in Activities of Daily Living (ADL)

The Index of Independence* in Activities of Daily Living is based on an evaluation of the functional independence or dependence of patients in bathing, dressing, going to toilet, transferring, continence, and feeding.

A – independent in feeding, continence, transferring, going to toilet, dressing and bathing.

B – independent in all but one of these functions.

C – independent in all but bathing and one additional function.

D – independent in all but bathing, dressing and one additional function.

E – independent in all but bathing, dressing, going to toilet and one additional function.

F – independent in all but bathing, dressing, going to toilet, transferring and one additional function.

G – dependent in all six functions.

Other – dependent in at least two functions, but not classifiable as C, D, E or F.

*Independence means without supervision, direction, or active personal assistance, except as specifically noted. This is based on actual status and not on ability. A patient who refuses to perform a function is considered as not performing the function, even though he is deemed able.

Source: Katz and Akpom, 1976, Figure 1, p. 496.

Table 7). In the study by Applegate and colleagues, over half of the patients were classified as completely independent before surgery; for these patients, the scale could not identify further improvement.

This emphasizes the need to find measures that can discriminate among groups with very different visual characteristics—a need that led Bernth-Petersen (1981) to construct a specific instrument. But there is an equal need for health status measurement instruments to be reliable and well-validated on a range of different subject groups so as to permit comparisons among quite different interventions.

Fortunately, there now exists a range of general instruments that may meet both these requirements; unfortunately, it appears they have not as yet been tested in use with visually impaired people.

For example, the Sickness Impact Profile (Bergner and coworkers, 1976 and 1981) is a wide-ranging instrument for patient self-completion, with 136 items covering 12 dimensions (see Table 8) but no explicit questions on visual functioning. However, certain questions relating to, for example, communications, mobility, work, and pastimes might identify problems of visual functioning. [An Anglicized version of the SIP—the Functional Limitations Profile—exists for use in the United Kingdom (Patrick, 1980)]. In each case, patients are asked to say whether particular statements (all indicating some form of restriction, limitation, or distress) apply to themselves and whether or not they are due to their health. The SIP has been widely used, particularly in the U.S., and comparative data are available for an increasing range of patient groups. Weights have been calculated for questions within each dimension, based on the public's perceptions of the relative importance of each item, and the results are normally presented as scores for each dimension. These scores can be aggregated to an overall SIP score, although this assumes that each dimension is of equal importance.

A similar but rather shorter instrument has been developed in the United Kingdom—the Nottingham Health Profile (Hunt, 1986). This has a similar multidimensional characteristic (see Table 9) and has been widely used in such diverse situations as the evaluation of heart transplantation (Buxton and coworkers, 1985) and

**Table 8**
**Sickness Impact Profile**

| Dimensions | Number of Statements |
|---|---|
| Sleep and rest | 7 |
| Work | 9 |
| Eating | 9 |
| Mobility | 10 |
| Home management | 10 |
| Recreation | 8 |
| Body care and movement | 23 |
| Ambulation | 12 |
| Social interaction | 20 |
| Alertness behaviour | 10 |
| Emotional behaviour | 9 |
| Communication | 9 |
| | 136 |

Source: Bergner et al., 1981.

population surveys (Curtis, 1985). The instrument devisers argue that it should be used only as a profile, again with weights available for each statement within dimensions, but not for weighting the dimensions themselves. In some studies, simple accounts of the numbers of statements affirmed have been used as a summary statistic. Again, this instrument has no questions relating directly to visual impairment but such limitations may well be reflected in responses to questions on physical mobility, emotional reactions, and social isolation, for example. Only by application to visually impaired populations can its sensitivity to the effects of sight problems be assessed.

The McMaster Health Index Questionnaire (MHIQ) (Chambers and coworkers, 1982) is a more complex and varied instrument covering the three main areas of physical function, emotional function, and social function. It consists of 59 items and involves a number of contingent questions; answers to some questions are rated on a scale 1-3 or 1-5. Within each main area, the responses to items are dichotomized and weights of "1" for a good health response and "0" for a poor health response are assigned to each item and added to calculate an MHIQ physical function score, an MHIQ social function score, and an MHIQ emotional function score. Unlike the SIP and NHP, the MHIQ includes a specific question on sight—a contingent question involving a simple rating procedure (see Table 10).

A number of other such instruments are available, or are being developed, and they carry with them a large detailed literature. Some of these have been developed for particular major studies—for example, the Rand Health Insurance Study (Ware and coworkers, 1984). To date, there has been only a limited amount of empirical comparative analysis of the results from using different instruments on the same subjects. The trial of auranofin therapy for patients with rheumatoid arthritis is a noticeable exception (Bombardier and coworkers, 1986). The study included 20 nontraditional measures of functional performance, pain, global impression, and utility. Until more comparative studies of this type have been completed, the choice of such instruments must reflect views as to their reliability and validity, likely relevance and discriminatory power in a particular context, and, not least of all, practicality (Ware and coworkers, 1981).

## Measuring Utility of States Of Visual Impairment

In the context of economic evaluation, the concept of the utility of health states is crucial. While many profiles and health indices may give a very clear picture of the health state of the individuals or groups in question and show how health states differ between groups, they generally do not provide an indication of how the resultant state is valued. Utility is a measure of "the value or worth of a specific level of health status (or improvement in health status) and can be measured by the preferences of individuals or society for any particular set of health outcomes" (Drummond, 1987).

Thus, it is not sufficient for the purposes of economic analysis simply to be able to systematically describe or measure health status. Additionally, a valuation process is required. These valuations will differ according to individual or group perceptions of what aspects of health or its

## Table 9
## Nottingham Health Profile: Dimensions, Statements and Weights

| | |
|---|---|
| PHYSICAL MOBILITY | |
| I find it hard to reach for things | 9.30 |
| I find it hard to bend | 10.57 |
| I have trouble getting up and down stairs or steps | 10.79 |
| I find it hard to stand for long (e.g., at the kitchen sink, waiting for a bus) | 11.20 |
| I can only walk about indoors | 11.54 |
| I find it hard to dress myself | 12.61 |
| I need help to walk about outside (e.g., a walking aid or someone to support me) | 12.69 |
| I'm unable to walk at all | 21.30 |
| | 100.00 |
| PAIN | |
| I'm in pain when going up and down stairs or steps | 5.83 |
| I'm in pain when I'm standing | 8.96 |
| I find it painful to change position | 9.99 |
| I'm in pain when I'm sitting | 10.49 |
| I'm in pain when I walk | 11.22 |
| I have pain at night | 12.91 |
| I have unbearable pain | 19.74 |
| I'm in constant pain | 20.86 |
| | 100.00 |
| SLEEP | |
| I'm waking up in the early hours of the morning | 12.57 |
| It takes me a long time to get to sleep | 16.10 |
| I sleep badly at night | 21.70 |
| I take tablets to help me sleep | 22.37 |
| I lie awake for most of the night | 27.26 |
| | 100.00 |
| ENERGY | |
| I soon run out of energy | 24.00 |
| Everything is an effort | 36.80 |
| I am tired all the time | 39.20 |
| | 100.00 |
| SOCIAL ISOLATION | |
| I'm finding it hard to get on with people | 15.97 |
| I'm finding it hard to make contact with people | 19.36 |
| I feel there is nobody I am close to | 20.13 |
| I feel lonely | 22.01 |
| I feel I am a burden to people | 22.53 |
| | 100.00 |
| EMOTIONAL REACTIONS | |
| The days seem to drag | 7.08 |
| I'm feeling on edge | 7.22 |
| I have forgotten what it is like to enjoy myself | 9.31 |
| I lose my temper easily these days | 9.76 |
| Things are getting me down | 10.47 |
| I wake up feeling depressed | 12.01 |
| Worry is keeping me awake at night | 13.95 |
| I feel as if I'm losing control | 13.99 |
| I feel that life is not worth living | 16.21 |
| | 100.00 |

Derived from: Hunt, McEwen, and McKenna, 1986.

## Table 10
## McMaster Health Index Questionnaire: Item Relating to Sight

8. Do you wear glasses?

1 NO → c. Do you have any trouble seeing ordinary newsprint?

1 NEVER
2 SOMETIMES
3 ALWAYS

d. Do you have a headache after watching television or reading?

1 NEVER
2 SOMETIMES
3 ALWAYS

2 YES → a. Do you have any trouble seeing ordinary newsprint when you wear your glasses?

1 NEVER
2 SOMETIMES
3 ALWAYS

b. Do you have a headache after watching television or reading when you wear your glasses?

1 NEVER
2 SOMETIMES
3 ALWAYS

Source: Chambers et al., 1982, p. 343.

## Table 11
## The Rosser Classification and Valuation Matrix

| Disability Rating | | Distress Rating | | | |
|---|---|---|---|---|---|
| | | A. No Distress | B. Mild | C. Moderate | D. Severe |
| I | No disability | 1.000 | 0.995 | 0.990 | 0.967 |
| II | Slight social disability | 0.990 | 0.986 | 0.973 | 0.932 |
| III | Severe social disability and/or slight physical impairment | 0.980 | 0.972 | 0.956 | 0.912 |
| IV | Physical ability severely limited (e.g. light housework only) | 0.964 | 0.956 | 0.942 | 0.870 |
| V | Unable to take paid employment or education, largely housebound | 0.946 | 0.935 | 0.900 | 0.700 |
| VI | Confined to chair or wheelchair | 0.875 | 0.845 | 0.680 | 0.000 |
| VII | Confined to bed | 0.677 | 0.564 | 0.000 | – 1.486 |
| VIII | Unconscious | – 1.028 | * | * | * |

Scale defined so that: Healthy = 1.0; Dead = 0.0.

* Denotes invalid combination of disability and distress.

Source: Kind, Rosser, and Williams, 1982, p. 160.

absence are most important. Despite the individuality of values, research has shown a considerable degree of consistency of values among different groups (Torrance, 1986; Kind and coworkers, 1982). In terms of public policy decisions, it is debatable whether the values should necessarily be those of the individuals affected, or whether the appropriate values should be those of informed members of the general public or community representatives (Drummond, 1987).

Utility values need not, and typically do not, take the form of monetary values, although monetary values derived from willingness-to-pay approaches have been calculated (Bombardier, 1986). The preferred method is to obtain valuations for health states relative to each other and to "full health" and "death," so that the utilities can be used in combination with morbidity and mortality data to calculate the quality-adjusted life-years (QALYs) gained. This then provides an overall value measure of outcome from particular health care interventions.

The Index of Well-Being (IWB) (Kaplan and coworkers, 1976) provides both a classification system for description and estimates of the associated values. The classification system involves four attributes: mobility, physical function, and social function (common to many instruments) plus a "symptom-problem" complex. The first three dimensions produce 43 function levels, each associated with an estimated utility or level of well-being; these are then adjusted to take account of 36 symptom-problem complexes which affect the utility value. These include just two relating directly to sight:

**Cl:** Any trouble seeing—including wearing glasses or contact lenses.

**C2:** Pain or discomfort in one or both eyes, such as burning or itching.

Utility values have been derived from population surveys. These indicate that the presence of symptom/problem Cl or C2 would reduce the utility by approximately 0.238 or 0.223, respectively.

Rosser and colleagues (Rosser, 1978) have developed simpler classification focusing on two dimensions of disability and distress, producing a matrix of 29 possible states (Table 11). The scale of utility values for these states was obtained by the method of magnitude estimation in interviews with 70 subjects, and standardized by assigning a value of 1 to freedom from disability and distress and to death (Kind and coworkers, 1982).

**Table 12**
**TTO Health State Utilities From Sackett and Torrance**

| Health State Scenarios | Duration | | |
|---|---|---|---|
| | 3 Months | 8 Years | Life |
| Depression | 0.44 | — | — |
| Home confinement for tuberculosis | 0.68 | — | — |
| Home confinement for unnamed contagious disease | 0.65 | — | — |
| Hospital confinement for tuberculosis | 0.60 | — | — |
| Hospital dialysis | 0.62 | 0.56 | 0.32 |
| Hospital confinement for unnamed contagious disease | 0.56 | 0.33 | 0.16 |
| Home dialysis | — | 0.65 | 0.40 |
| Kidney transplant | — | 0.58 | — |
| Mastectomy for breast cancer | — | 0.48 | — |
| Mastectomy for injury | — | 0.63 | — |

Source: Sackett and Torrance, 1978, Table 2.

Torrance and colleagues (e.g., Torrance, 1976; Sackett, 1978) have used a different approach. Rather than formal (and constricting) classifications, they used a narrative style in "scenarios" describing the health states in question. Table 12 presents the utility values (Sackett and Torrance, 1978) derived from a sample of 246 members of the local general population using a stratified sampling frame. Results showed that the utility of a unit of time in a health state was a function of the time spent in that state and decreases as the length of time in the health state is extended. The estimation procedure was that of the "time trade-off" (TTO); for a review of that approach, see Buxton and Ashby (forthcoming).

No such TTO estimations have been published relating to broad scenarios of states of visual impairment. However, following earlier work applying TTO to scenarios relating to women treated for breast cancer (Buxton and coworkers, 1987), a study conducted during 1987 at Brunel University includes a pilot visual-impairment scenario (see Table 13). Results from this study should be available early in 1988. It will report the utility values of about 80 medical professionals and lay subjects for this and five other quite different health-state scenarios.

**Table 13**
**TTO Scenario As Used**
**In Current Research at Brunel**

| Partially Sighted |
|---|
| You take more time around the house to perform basic tasks such as cooking and housework. You can manage stairs or steps without much difficulty. |
| Going outdoors is not a problem even if a walking aid is required. You enjoy the social contact this provides. Sometimes you are irritated when people want to help more than you think is necessary, but overall you enjoy a high degree of independence. |
| Occasionally you feel depressed and anxious about the possibility of further sight deterioration. On the whole, however, you enjoy life and appreciate your ability to perceive the environment, scenery and so on. |
| You have several interesting hobbies and are able to do some gardening at home. You have had to give up your old job but now work part-time after specialised vocational training. You enjoy the increased leisure-time and feel that life is less stressful as a result. |

The basic concept of the utility of visual impairment has been explored by Card and colleagues (1977). He used a technique conceptually allied to the TTO, that of the "standard gamble," to elicit utility values from eight subjects for a series of scenarios relating to typical patients with differing degrees of visual acuity.

Table 14 presents one of these scenarios, for which the utility values from the eight subjects interviewed range from 0.805 to 0.989. However, the investigators' focus was exclusively on the single dimension of visual impairment. Therefore, the utility scale was constructed so that the lower utility bound of 0 corresponded to a distance visual acuity of no perception of light, and the upper bound of 1 to a distance visual acuity of 6/6. Thus, the whole scale would represent but part of the other utility scales described above, which conventionally set the lower bound 0 as dead, and the upper bound 1 as perfect health. In Card's scenarios, utility is in effect solely a function of visual acuity and the two have a log-linear relationship.

**Table 14**
**Visual Impairment Scenario**

| Health State E | | |
|---|---|---|
| VA (distance) | : | 6/60 |
| VA (near) | : | N.24 |
| Occupations | : | On partially sighted register. Telephone operator and light engineering. Could do unskilled labouring, but at a lower wage. Wife would work. |
| Recreations | : | Some gardening possible. Varied walks possible. Could cook and do shopping (not too complicated). |

Source: Card et al., 1977, p. 170.

A further approach to utility measurement has combined the approach of a formal descriptive classification of states (rather than narrative scenarios) with the use of TTO (Torrance and coworkers, 1982). In this case, the utility of a health state can be determined by placing it appropriately into a multiattribute classification for which values have previously been elicited, rather than specifically eliciting values for "unique" scenarios. The approach has been described thus:

> "If the health states are described by a classification system of the type appropriate for multi-attribute utility theory methods, . . . an appropriate multi-attribute utility model (additive, multiplicative, etc.) must be determined, and this in turn specifies the utility measurement which needs to be made . . . However, the fundamental utility measurements are then made using the basic measurement techniques . . . Thus, the multi-attribute utility theory approach does not introduce any new methods of measuring preferences. Rather, it introduces a way of selecting specific preferences to be measured and combining them into a mathematical model of the subject's utility structure. The model can then be used to determine the utility values for all possible health states in the classification system." (Torrance, 1986, p. 25)

Torrance's multiattribute utility (MAU) model involves a four-dimensional classification as in Table 15. The multiplicative utility factors are shown in Table 16. One of the specific health problems in the fourth dimension of MAU is described as "having trouble seeing even when wearing glasses."

Using this MAU model, Drummond (1987) calculates that "an emotionally well-adjusted blind person would have a health state utility of 0.48, a poorly adjusted person one of 0.35." This, he argues, suggests that the potential gain in utility from completely restoring sight is in the order of 0.6. Such estimates must at this stage be seen as tentative and exploratory; further utility assessment focusing on visual impairment is needed before such estimates can be used with conviction. The importance of doing such work lies in the increasing use now being made of such utility measurement in the calculation of QALYs as a standard unit of output from health services. This approach is favored by many health economists. The basic tenet is that potential patient benefits from treatment are dichotomous: both quality and length of life should be considered. There is an implicit tradeoff in the outcome of health care interventions between quality of life (QOL) and life expectancy, which needs to be made explicit for rational decision-making purposes. Some treatments (e.g., hip replacement) may improve QOL immensely without substantially increasing length of life, while other interventions (e.g., renal dialysis) may significantly prolong life without greatly improving quality.

The QALY then is a standardized unit of outcome of health care interventions, and the basic objective is to maximize the number of QALYs obtained from the available budget. The utility values elicited by the various measurement techniques form the basis of the QOL calculation, and estimates of expected survival (additional to prognosis without treatment) must also be made. Additional years of life expectancy are therefore "scaled down' by the QOL adjustment factor.

Thus, "cost per QALY" can be used as a measure of relative value obtained from increasingly scarce health care resources. In the Canadian context, "a cost per QALY league table" has been put together by Torrance (1986) on the basis of the results of a number of studies of different patient groups and therapies (Table 17). In the United Kingdom, the same conceptual approach (but based on the Rosser utility values) has been widely advocated by Williams (e.g., Williams, 1985). The approach has now entered the policy debate in the context of the relative cost-utility of breast cancer screening (DHSS, 1986) (Table 18). In addition, Drummond (1987) estimated the cost per QALY for cataract operations to be approximately £500, although his methodology was not strictly comparable to that of Williams (1985).

At the present time, the exact values produced by the application of the cost-per-QALY concept needs to be treated with caution, but it seems likely that this framework will be used increasingly in the policy arena to illuminate the problem of establishing relative priorities in spending, among the numerous claims from quite different health care programs.

## Particular Problems in Using Existing Measurement Techniques With Visually Impaired People

Compared with some of the *conceptual* and *methodological* problems raised in attempting to measure and value states of health for patients or members of the general population, the additional *practical* problems in measurement or

## Table 15
## Multi-Attribute Utility Health State Classification

**$X_1$ Physical Function: Mobility and Physical Activity[a]**

| Level $x_1$ | Code | Description |
|---|---|---|
| 1 | P1 | Being able to get around the house, yard, neighbourhood or community WITHOUT HELP from another person; AND having NO limitation in physical ability to lift, walk, run, jump or bend. |
| 2 | P2 | Being able to get around the house, yard, neighbourhood or community WITHOUT HELP from another person; AND having SOME limitation in physical ability to lift, walk, run, jump or bend. |
| 3 | P3 | Being able to get around the house, yard, neighbourhood or community WITHOUT HELP from another person; AND NEEDING mechanical aids to walk or get around. |
| 4 | P4 | NEEDING HELP from another person in order to get around the house, yard, neighbourhood or community; AND having SOME limitations in physical ability to lift, walk, run, jump or bend. |
| 5 | P5 | NEEDING HELP from another person in order to get around the house, yard, neighbourhood or community; AND NEEDING mechanical aids to walk or get around. |
| 6 | P6 | NEEDING HELP from another person in order to get around the house, yard, neighbourhood or community; AND NOT being able to use or control the arms and legs. |

**$X_2$ Role Function: Self-Care and Role Activity[a]**

| Level $x_2$ | Code | Description |
|---|---|---|
| 1 | R1 | Being able to eat, dress, bathe, and go to the toilet WITHOUT HELP; AND having NO limitations when playing, going to school, working or in other activities. |
| 2 | R2 | Being able to eat, dress, bathe, and go to the toilet WITHOUT HELP; AND having SOME limitations when playing, going to school, working or in other activities. |
| 3 | R3 | Being able to eat, dress, bathe, and go to the toilet WITHOUT HELP; AND NOT being able to play, go to school, or work. |
| 4 | R4 | NEEDING HELP to eat, dress, bathe or go to the toilet, AND having SOME limitations when playing, going to school, working or in other activities. |
| 5 | R5 | NEEDING HELP to eat, dress, bathe or go to the toilet, AND NOT being able to play, attend school or work. |

[a]Multiple choices within each description are applied to individuals as appropriate for their age. For example, a 3-year-old child is not expected to be able to get around the community without help from another person.

## [Table 15 contd.]

**$X_3$ Social-Emotional Function: Emotional Well-Being and Social Activity**

| Level $x_3$ | Code | Description |
|---|---|---|
| 1 | S1 | Being happy and relaxed most or all of the time, AND having an average number of friends and contacts with others. |
| 2 | S2 | Being happy and relaxed most or all of the time, AND having very few friends and little contact with others. |
| 3 | S3 | Being anxious and depressed some or a good bit of the time, AND having an average number of friends and contacts with others. |
| 4 | S4 | Being anxious and depressed some or a good bit of the time, AND having very few friends and little contact with others. |

**$X_4$ Health Problem[b]**

| Level $x_4$ | Code | Description |
|---|---|---|
| 1 | HI | Having no health problem. |
| 2 | H2 | Having a minor physical deformity or disfigurement such as scars on the face. |
| 3 | H3 | Needing a hearing aid. |
| 4 | H4 | Having a medical problem which causes pain or discomfort for a few days in a row every two months. |
| 5 | H5 | Needing to go to a special school because of trouble learning or remembering things. |
| 6 | H6 | Having trouble seeing things even when wearing glasses. |
| 7 | H7 | Having trouble being understood by others. |
| 8 | H8 | Being blind OR deaf OR not able to speak. |

[b]Individuals with more than one health problem are classified according to the problem they consider the most serious.

Source: Drummond, Stoddart, and Torrance, 1987, Table 6.1.

valuation arising from the visual impairment of subjects are likely to be minor.

Most self-assessment instruments take the form of questionnaires. In many cases, these would simply need to be presented in a way that accommodated those with poor reading vision. For some totally blind people, Braille versions would be appropriate. For others, a companion or spouse might normally read documents for the visually handicapped person and record his or her responses. However, this may introduce a contaminating factor with the second person influencing, intentionally or unintentionally, the responses. This would obviously be undesirable and could perhaps be avoided by using audiotapes or an interviewer. The practicality of the

**Table 16**
**MAU Classification: Multiplicative Utility Factors**

| Physical Function | | Role Function | | Socio-Emotional Function | | Health Problem | |
|---|---|---|---|---|---|---|---|
| Level | Multiplicative Utility Factor $m_1$ | Level | Multiplicative Utility Factor $m_2$ | Level | Multiplicative Utility Factor $m_3$ | Level | Multiplicative Utility Factor $m_4$ |
| P1 | 1.00 | R1 | 1.00 | S1 | 1.00 | H1 | 1.00 |
| P2 | 0.91 | R2 | 0.94 | S2 | 0.96 | H2 | 0.92 |
| P3 | 0.81 | R3 | 0.77 | S3 | 0.86 | H3 | 0.91 |
| P4 | 0.80 | R4 | 0.75 | S4 | 0.77 | H4 | 0.91 |
| P5 | 0.61 | R5 | 0.50 | | | H5 | 0.86 |
| P6 | 0.52 | | | | | H6 | 0.84 |
| | | | | | | H7 | 0.83 |
| | | | | | | H8 | 0.74 |

Source: Drummond, Stoddart, and Torrance, 1987, Table 6.2.

latter depends on the scale and geographical extent of the sample involved and the research resources available.

Similar problems would arise with the use of scenarios in utility measurement. In this case, audiotapes would have the advantage of permitting repeated playback if required. The interview "props" often used with TTO (the visual sliding-scale boards) or with Standard Gamble (visual probability discs) would similarly need to be adapted for the partially sighted or, more probably, replaced altogether to ensure that the presentation to all subjects is identical. This is important given the concerns about the way in which the framing of questions or scenarios may influence responses (Llewellyn-Thomas, 1984).

In obtaining utility valuations, it is currently a matter of debate whether utilities should be obtained from those directly affected in this case, from visually impaired people themselves or whether utilities should be elicited from society as a whole—from professional or elected decisionmakers or from some other pertinent group. The choice of source of such values is essentially political. At this stage in the research it is important to establish whether significant and systematic differences in values exist among different groups. Few major differences have so far been identified (Torrance, 1986), although Sackett and Torrance (1978) found that dialysis patients put a higher utility on a dialysis health state scenario than those who had no direct experience of it.

Generally, it would seem reasonable to predict that the visual impairment of potential subjects is a factor that will influence, and to some extent constrain, the presentation of measurement and valuation instruments. However, it should not present any insuperable difficulties.

## Conclusions and Recommendations

The review has shown that while there has been some measurement and valuation of health states relating to visual impairment, this is so far limited. In particular, most clinical trials of interventions to reduce visual impairment restrict their measurement of success to traditional indicators

## Table 17
## Comparative Cost-Utility Results for Selected Programs[a]

| Program | Cost/QALY[b] Gained In US $ 1983 |
|---|---|
| PKU screening | < 0 |
| Post-partum anti-D | < 0 |
| Ante-partum anti-D | 1,220 |
| Coronary artery bypass grafting for left main coronary artery disease | 4,200 |
| Neonatal intensive care, 1000-1499 g | 4,500 |
| T4 (thyroid) screening | 6,300 |
| Treatment of severe hypertension (diastolic ≥ 105mm Hg) in males age 40 | 9,400 |
| Treatment of mild hypertension (diastolic 95-104 mm Hg) in males age 40 | 19,100 |
| Estrogen therapy for menopausal symptoms in women without a prior hysterectomy | 27,000 |
| Neonatal intensive care, 500-999 g | 31,800 |
| Coronary artery bypass surgery for single vessel disease with moderately severe angina | 36,300 |
| School tuberculin testing program | 43,700 |
| Continuous ambulatory peritoneal dialysis | 47,100 |
| Hospital haemodialysis | 54,000 |

[a]These studies use similar, but not identical, methods. Generally, costs are net health care costs; however, discount rates and preference weights are not completely consistent. Differences in methods should be considered when comparing the relative cost-utility.

[b]Quality-adjusted life-year.

Source: Torrance, 1986, Table 1, p. 6.

such as visual acuity or visual field. However, in other branches of medicine there is a growing interest in assessing the ways in which therapy improves the quality of life and several general scales have been developed and validated. The relative lack of progress in the eye care field is surprising, since the major impact of interventions is on the quality of life, not length of life. Given the increased emphasis on value from the use of health care resources, it is important that measurable results are produced and expressed in a way that enables comparisons of cost and outcome to be made with those in other clinical fields.

In order to meet this need, the following recommendations are made:

1. In selected clinical trials, one or more existing general instruments for measuring quality of life should be used alongside traditional clinical outcome measures. This will enable assessment of the sensitivity of the general

**Table 18**
**Comparison of Costs Per Quality-Adjusted Life-Year (QALY) Gained for Various Health Care Procedures**

| Procedure | Value of the Extra Costs Per QALY (£1983/84 Prices) |
|---|---|
| Pacemaker implantation for atrioventricular heart block | 700 |
| Hip replacement | 750 |
| Coronary artery bypass grafting (CABG) for main vessel disease | 1,040 |
| Kidney transplantation | 3,000 |
| | Breast cancer screening estimated by working group to fall in this range |
| Heart transplantation | 5,000 |
| CABG for moderate angina with one vessel disease | 12,000 |
| Hospital haemodialysis | 14,000 |

Source: DHSS, 1986, Table 9.2, p. 62.
[based on Williams, A. (1985), Economics of coronary artery bypass grafting, *British Medical Journal* 291: 362-9, 1985, Tables IV and V]

instruments and of the convergent validity between different measures, as well as exploration of the relationship between visual acuity, patient function, and quality of life.

2. As a parallel exercise, a number of initiatives should be launched to consider the content appropriate to a specific quality-of-life measurement instrument suitable for use in evaluating eye care interventions:
   - A criterion group should be assembled to discuss which aspects of quality of life are likely to be affected by visual impairment.
   - Patients with particular eye problems should be surveyed to assess how their condition affects their quality of life.
3. In the light of the results of (1) and (2) above, a decision should be made on whether:
   - An existing general index is adequate for these purposes.
   - Relevant extension can be made to an existing instrument in order to sensitize it for use in the eye care field.
   - Sub-sections of general health indices that are particularly relevant to the impact of visual impairment should be incorporated in a new composite health index.
   - A completely new index should be developed and validated for incorporation into future clinical trials.

4. In order to compare the relative value for money spent on interventions in eye health care with those in other health fields, utility values should be calculated for key states of visual impairment. The choice of health states for utility assessment could either be:
   - Opportunistic, in that measurements could be made alongside clinical trials in the eye care field where the clinical researchers concerned are willing to cooperate.
   - Based on the relative importance of the eye diseases (in terms of their prevalence and/or size of economic burden) that the health states represent.

Assessments of utility values for states of visual impairment should be made both on individuals suffering from particular eye diseases, where valuations can be based on actual experience, and on the general public, where valuations can be directly compared with those for a wide range of health states.

The National Eye Institute and other research-funding agencies should take a lead in these initiatives by encouraging new research proposals for instrument development and health state valuation, and by ensuring that validated quality-of-life measures are added to some existing or proposed clinical research studies.

## References


Alexander MF et al. Assessment of visual function in patients with age-related macular degeneration and low visual acuity. *Investigative Ophthalmology and Visual Science* 27(suppl)3, 1986.

Applegate WB, Miller ST et al. Impact of cataract surgery with lens implantation on vision and physical function in elderly patients. *Journal of the American Medical Association* 257(8):1064-1066, 1987.

Ash DDG, Keegan DL, Greenough T. Factors in adjustment to blindness. *Canadian Journal of Ophthalmology* 13(1):15-21, 1978.

Bergner M et al. The sickness impact profile: Conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services* 6(3):393-415, 1976.

Bergner M et al. The sickness impact profile: Development and final revision of a health status measure. *Medical Care* 19:787-805, 1981.

Bernth-Petersen P. Visual functioning in cataract patients: Methods of measuring and results. *Acta Ophthalmologica* 59:198-205, 1981.

Bernth-Petersen P. Outcome of cataract surgery: Socio-economic aspects. *Acta Ophthalmologica* 60:461-8, 1982.

Bombardier C et al. Auranofin therapy and quality of life in patients with rheumatoid arthritis. *The American Journal of Medicine* 81:565-577, 1986.

British Multicentre Photocoagulation Trial. Photocoagulation in treatment of diabetic maculopathy. *The Lancet* 2:1110-1113, Dec. 6, 1975.

British Multicentre Photocoagulation Trial. Proliferative diabetic retinopathy: Treatment with xenon-arc photocoagulation. *British Medical Journal* 1:739-741, Mar. 19, 1977.

Burack B. Inter-disciplinary classification for the aged. *Journal of Chronic Diseases* 18:1059-1064, 1965.

Buxton MJ et al. Costs and Benefits of the Heart Transplant Programmes at Harefield and Papworth Hospitals London: D.H.S.S. Research Report No 12, H.M.S.0, 1985.

Buxton MJ, Ashby J, O'Hanlon M. A comparative analysis based on an empirical study using the time tradeoff approach in relation to health states one year after treatment for breast cancer. UK: Brunel University, Health Economics Research Group, July 1987.

Buxton MJ and Ashby J. The time tradeoff approach to health state valuation: A North American history and a UK application. *In*: Teeling-Smith G (ed) Measuring Health: A Practical Guide, J. Wiley (forthcoming).

Calek O. Distressful life situations and the emotional strain felt by subjects with severe visual defects. *American Foundation for the Blind Research Bulletin* No 25, 59-67, 1973.

Card WI, Rusinkiewicz M, Phillips CI. Utility estimation of a set of states of health. *Methods of Information in Medicine* 16(3):168-175, 1977.

Chambers LW et al. The McMaster health index questionnaire as a measure of the quality of life for patients with rheumatoid disease. *Journal of Rheumatology* 9:780-4, 1982.

Chambers LW et al. Sensitivity to change and the effect of mode of administration on health status measurement. *Medical Care* 25(6): 470-480, 1987.

Colenbrander A. Dimensions of visual performance. *Transactions of the American Academy of Ophthalmology and Otolaryngology* 83:347-350, 1977.

Cullinan TR. The Epidemiology of Visual Disability: Studies of Visually Disabled People in the Community, Health Services Research Unit, University of Kent at Canterbury, UK, Report No. 28, 1977.

Curtis S. Intra-Urban Variations in Health Care, vol 1. Department of Geography and Earth Science, Queen Mary College, London University, London, 1985.

Department of Health and Social Security. Breast Cancer Screening: Report of a Working Party Chaired by Professor Sir Patrick Forrest London: H.M.S.0., 1986.

Diabetic Retinopathy Study Research Group. Preliminary report on effects of photocoagulation therapy. *American Journal of Ophthalmology* 81(4):383-396, 1976.

Diabetic Retinopathy Study Research Group. Assessing possible late treatment effects in stopping a clinical trial

early: A case study. *Controlled Clinical Trials* 5:373-381, 1984.

Drummond MF. Economic Aspects of Cataract, Consultant's Report for the National Eye Institute, Bethesda, MD, 1987.

Drummond MF, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes, Oxford University Press, 1987.

Early Treatment Diabetic Retinopathy Study Research Group. Photocoagulation for diabetic macular edema. *Archives of Ophthalmology* 103:1796-1806, 1985.

Ferris FL et al. New visual acuity charts for clinical research. *American Journal of Ophthalmology* 94:91-96, 1982.

Genensky SM. Acuity measurements - Do they indicate how well a partially sighted person functions or could function? *American Journal Optometry and Physiological Optics* 53(12):809-812, 1982.

Haase KW and Bryant EE. Development of a scale designed to measure functional distance vision loss using an interview technique. *Proceedings of the American Statistical Association*, Social Statistics Section 274-279, 1973.

Hunt SM, McEwen J, McKenna SP. Measuring Health Status. London: Croom Helm, 1986.

Jastrzembska ZS. Social and psychological aspects of blindness: A sampling of the literature. *American Foundation for the Blind Research Bulletin* 25:169-173, 1973.

Kaplan RM, Bush JW, Berry CC. Health status: Types of validity and the index of well-being. *Health Services Research* 11:478-507, 1976.

Katz S and Akpom CA. A measure of primary sociobiological functions. *International Journal of Health Services* 6(3):493-507, 1976.

Kind P, Rosser RW, Williams A. Valuation of quality of life: Some psychometric evidence. *In*: Jones-Lee MW, (ed) The Value of Life and Safety, North-Holland Publishing Co., 1982.

Llewellyn-Thomas H et al. Describing health states: Methodological issues in obtaining values for health states. *Medical Care* 22(6):543-552, 1984.

Macular Photocoagulation Study Group. Argon laser photocoagulation for neovascular maculopathy: Three-year results from randomised controlled trials. *Archives of Ophthalmology* 104:694-701, 1986.

Oehler-Giarratana J. An Exploratory Study of Psychological Reactions to Visual Loss and Blindness in Patients with Diabetic Retinopathy, Ph.D. thesis, Boston: Boston University, 1980.

Patrick D. Standardisation of comparative health status measures: Using scales developed in America in an English-speaking country. *In*: Proceedings of the Third Biennial Conference on Health Survey Research Methods, National Center for Health Services Research, U.S. Department of Health and Human Services, Publication No PHS 81-3268, 216-220, 1980.

Pfeiffer E, Johnson TM, Chiofolo RC. Functional assessment of elderly subjects in four service settings. *Journal of the American Geriatric Society* 29:434-437, 1981.

Rosser RW and Kind P. A scale of valuations of states of illness: Is there a social consensus? *International Journal of Epidemiology* 7(4):347-358, 1978.

Sackett DL and Torrance GW. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases* 31(11):697-704, 1978.

Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences* 10:129-136, 1976.

Torrance GW. Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics* 5:1-30, 1986.

Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Operations Research* 30(6):1043-1069, 1982.

Ware JE et al. Choosing measures of health status for individuals in general populations. A*merican Journal of Public Health* 71(6):620-625, 1981.

Ware JE et al. Health status and the use of outpatient mental health services. *American Psychologist* 39:1090-1100, 1984.

Westheimer G. Scaling of visual acuity measurements. *Archives of Ophthalmology* 97:327-330, 1979.

Williams A. Cost-effectiveness of coronary artery bypass grafting: An economic appraisal. *British Medical Journal* 291:326-9, 1985.

Wulsin LR, Jacobson AM, Rand LI. Psychosocial aspects of diabetic retinopathy. *Diabetes Care* 10(3)367-373, 1987.

# Appendix B.
# Selective Annotated Bibliography Of Clinical Studies and Surveys of the Effects Of Eye Disease or Its Treatment

## A

**Applegate, W.B. and Miller, S.T.** et al. [1987]. "Impact of Cataract Surgery with Lens Implantation on Vision and Physical Function in Elderly Patients," *Journal of the American Medical Association*, Vol. 257, No. 8, pp. 1064-1066.

Prospective study of 293 elderly patients undergoing cataract surgery with intraocular lens implantation to determine the impact of surgery on vision and on subjective and objective measures of patient function. The subjective measures included activities of daily living (ADL) and self-report of vision-dependent activities, and objective measures included timed manual performance tests and mental status tests.

Visual acuity in the surgical eye improved from a mean of 20/200 before surgery to 20/40 four months after surgery; improvement was maintained at one year. Positive changes occurred in some subjective measures of patient function, such as reported ADL and patient report of vision-dependent activities, but these changes were modest. The most marked changes in patient function occurred in the objective measures. Mental status improved significantly at one year, while timed manual performance improved significantly at four months and one year.

**Ash, D.D.G., Keegan, D.L. and Greenough, T.** [1978] "Factors in Adjustment to Blindness," *Canadian Journal of Ophthalmology*, Vol. 13(1), pp. 15-21.

Authors wished to explore the relationship between certain blinding eye diseases (e.g., diabetic retinopathy, glaucoma), residual vision and psychosocial adjustment, including acceptance of blindness. 114 patients fulfilled the criteria for blindness and availability for interview. Structured interview used to collect demographic and historical information from subjects regarding four major phases in the development of visual loss. Also recorded were the time of greatest psychological disturbance, ability to cope and the social effects on aspects of the subjects' lives. Medical and ophthalmological assessment carried out.

Main measures of psychological adjustment were 3 scales of the Minnesota Multiphasic Personality Inventory (MMPI) which included 330 questions measuring clinical symptoms of anxiety, depression and somatic psychological complaints.

Social adjustment was measured by an 86-item questionnaire modified for blindness from the Gunzberg Progress Assessment Chart of Social Development. Sub-scales assessing success in self-help, occupation, communication and socialisation were used as well as a global social adjustment total (SAT) based on all 86 items.

## B

**Bernth-Petersen, P.** [1981] "Visual Functioning in Cataract Patients: Methods of Measuring and Results," *Acta Ophthalmologica*, Vol. 59, pp. 198-205.

Recognises that in many patients with visual impairment it is difficult to equate visual acuity measurements with the patient's ability or inability of visual functioning. In evaluating benefits of treatment, we require standardised measurements of levels in visual functioning in order to supplement visual acuity. Presents an Index for measuring Visual Functioning on the basis of self-assessment. "Visual functioning" is defined as the performance of everyday tasks which are vision-dependent, such as coping with work, housework, self-care and mobility.

Construct validity and reliability of the Index are shown to be sufficient, and "theoretically this index could be applied to any group of patients with visual impairment." Argued that the Index is easy to use when interviewing the patient, and usefully "supplements the clinical examination with information about practical visual functioning and the social conditions." Classification according to visual function is possible by applying a specific scoring system to the Index.

**Bernth-Petersen, P.** [1982] "Outcome of Cataract Surgery. IV: Socio-Economic Aspects," *Acta Ophthalmologica*, Vol. 60, pp. 461-468.

Presents some occupational, social and economic aspects of cataract surgery evaluated on the basis of observations in a prospective data collection study. Outcome of cataract surgery assessed by measuring visual acuity in the operated eye at one-year follow-up, and by applying the Visual Functioning Index and obtaining patients' statements about quality of life ("better" or "unchanged"). This paper adds information on "patients' need for help from others and patients' occupational status pre-operatively and 1 year after cataract extraction." Occupational status a question simply of whether engaged in active employment; need for help involves a twofold classification of (a) receiving municipal domestic help and (b) dependent on help from spouse/family or friends.

**British Diabetic Association Committee on Blindness** [1967-69] A Report on Diabetic Blindness in the United Kingdom, Published by the British Diabetic Association.

Small-scale (N=16 subjects) social survey "to explore the experiences and attitudes of people who had been registered blind by reason of diabetic retinopathy." Several aspects examined, for example, degree of dependence (related to diabetes), general health levels, social circumstances, advanced warning of and adjustment to visual disability, financial circumstances and rehabilitation. Overall impression was that in addition to problems general to blind people, there were particular problems faced by blind people with diabetes. Some of the problems were physical, relating to the co-existence of diabetes and its accompanying complications—while others were emotional.

**British Multicentre Photocoagulation Trial** [1975] "Photocoagulation in Treatment of Diabetic Maculopathy," *The Lancet*, Dec. 6, 2, pp. 1110-1113.

**British Multicentre Photocoagulation Trial** [1977] "Proliferative Diabetic Retinopathy: Treatment with Xenon-arc Photocoagulation," *British Medical Journal*, Mar. 19, 1, pp. 739-741.

**British Multicentre Study Group** [1984] "Photocoagulation for Proliferative Diabetic Retinopathy: A Randomised Controlled Clinical Trial Using the Xenon-arc," *Diabetologia*, Vol. 26, pp. 109-115.

1975 article details a multicentre R.C.T. of xenon-arc photocoagulation for diabetic maculopathy. Interim report presenting visual acuity results of patients followed for up to three years. Clinical benefits discussed primarily in terms of visual acuity:

"To the patient, . . . retention of visual acuity is the most important index of successful treatment, and . . . this functional measure provides the most valid basis for therapeutic or prophylactic photocoagulation in diabetic retinopathy."

Article outlines method of coding visual acuity scores on a 10-point scale, where 6/5 or 6/6 = 1 and "no perception of light" = 10. [see also Townsend, C., Bailey, J and Kohner, E.M., 1979]

1977 article follows from first interim report in 1975; here reports the result of the multicentre study as it applies to proliferative diabetic retinopathy. Nothing more to add from 1975 report concerning benefits of treatment: again changes in visual acuity the measure of successful treatment. 1984 article presents final results of the randomised controlled study on patients with proliferative diabetic retinopathy, followed for at least 5 years and some for up to 7 years. Mean visual acuity scores compared to assess treatment effects over time.

**Burack, B.** [1965] "Inter-Disciplinary Classification for the Aged," *Journal of Chronic Diseases*, Vol.18, pp. 1059-1064.

Inter-disciplinary classification involving 4 categories: medical, functional capacity, projected goals and therapeutic category. Case illustrations are taken from residents at the Home for the Aged Blind at the Jewish Guild for the Blind in New York, e.g., Case No. 3, a mild diabetic with blindness secondary to cataracts and glaucoma.

The authors, however, "do not consider blindness as a major handicap influencing the ability to function." The classification is presented in the broadest possible terms and it is intended to be applicable to "all groups of disabled individuals of every age and health category." It is therefore not particularly helpful in assessing health status relating to specific conditions.

## C

**Calek, O.** [1973] "Distressful Life Situations and the Emotional Strain Felt by Subjects with Severe Visual Defects," *American Foundation for the Blind Research Bulletin*, No. 25, pp. 59-67.

Small-scale study of 11 subjects studying at the Secondary Music College for Youth with Visual Defects in Prague, 1970. All subjects fell in the lowest category in classification of visual impairment. Not a representative sample of all totally blind people.

Involves use of psychological test methods adapted for visually-impaired subjects. Rather complex techniques employed, although article is useful in identifying areas of emotional strain. The reason most often given by subjects (10 out of 11) for personal strain was the inability to achieve success in many activities at which visually-normal people succeed. The author reviewed different stress situations and then presented these to subjects in pairs: two analogous but slightly different situations were paired for comparison each time.

**Card, W.I., Rusinkiewicz, M. and Phillips, C.I.** [1977] "Utility Estimation of a Set of States of Health," *Methods of Information in Medicine*, Vol. 16, No. 3, pp. 168-175.

Very useful article which attempts to relate a clinical measure of seeing (visual acuity) with more general health status-type measures. Uses the economist's concept of utility, whereby attached to each particular health state there is an implied utility value. This value can be obtained by a number of different methods; the paper uses the Standard Gamble technique, the foundations of which rest on the axioms of "expected utility theory" [ref. Neumann, J. von and Morgenstern, O. : Theory of Games and Economic Behaviour, 3rd ed., Princeton University Press, 1953].

The subjects tested were four medical (including two ophthalmologists) and four non-medical subjects. "The set of states of health to be considered must refer to a particular individual and the way in which a given loss of vision affects his work and recreation must be specified."

The main finding of this article is that the utility function for each decision-maker was found to be linear against the logarithm of the visual acuity. The main problem is that the analysis involves states of health which are unidimensional, but the vast majority of states of health encountered in medical practice are multidimensional. Only further work along the lines of this paper could show if the analysis can satisfactorily be extended.

**Colenbrander, A.** [1977] "Dimensions of Visual Performance," *Transactions of the American Academy of Ophthalmology and Otolaryngology*, 83, pp. 347-350.

Points out that proposed guidelines from the "new" International Classification of Diseases classify levels of visual impairment using both visual acuity and visual field:

| | VA | | VF |
|---|---|---|---|
| No impairment | ≥ 20/25 | and | ≥ 120° |
| Slight impairment | < 20/25 | or | < 120° |
| Moderate impairment | < 20/60 | or | < 60° |
| Severe impairment | < 20/160 | or | ≥ 20° |
| Profound impairment | < cf at 3m | or | ≥ 10° |
| Near total impairment | < cf at 1m | or | ≥ 5° |
| Total impairment | No light perception | | |

Important because we should observe the "overall function of the visual organ," measuring varying degrees of impairment in terms of visual acuity, visual field, binocular vision, colour vision, or night vision.

**Cullinan, T.R.** [1977] The Epidemiology of Visual Disability: Studies of Visually Disabled People in the Community, Health Services Research Unit, University of Kent at Canterbury, UK, Report No. 28.

Comprehensive report in five parts:

Part I. Introduction.

Part II. Review of World Literature.

Part III. Visually Disabled People in Canterbury, 1974.

Appendix 3 of Part III contains a form asking "special questions for the visually disabled." Fairly generalised questionnaire which would not easily distinguish degrees of sight impairment other than between subjects who were totally blind and not totally blind.

Part IV. National Survey of Visually Disabled People, England and Wales, 1976-77. Part IV attempts, inter alia, to relate visual acuity scores to "social mobility"; e.g., 17 subjects with a visual acuity of 6/24 had "been out within the previous week," while only five subjects with a visual acuity of 3/60 and two subjects with a visual acuity of "no light perception" had done so.

Part V. Conclusions and Discussion. The Report provides three particular cautionary notes regarding the assessment of visual disability (Part V, p.140):

(a) no question, open or closed, directive or non-directive, can be used to estimate with any accuracy either distance or near visual acuity;

(b) the measurement of distance visual acuity is no accurate indication of near vision;

(c) the level of measured visual acuity, distance or near, provides no accurate guide to what can be achieved visually or to the extent of the handicap experienced—many can achieve with a little residual sight far more than others can do with considerably more.

"Nevertheless, if a system largely dependent on distance acuity is to continue to be used it seems logical to use a system such as the W.H.O. one, generous enough to include most of those with a near vision problem, rather than to exclude many of them as our current concepts of partial sight and blindness seem to do."

[Definition of Visual Impairment and Blindness, WHO 1973 given in Part II, Appendix 1].

# D

**Diabetic Retinopathy Study Research Group** (DRSRG) [1976] "Preliminary Report on Effects of Photocoagulation Therapy," *American Journal of Ophthalmology*, Vol. 81, No. 4, pp. 383-396.

**DRSRG** [1978] "Photocoagulation Treatment of Proliferative Diabetic Retinopathy: The Second Report of Diabetic Retinopathy Study Findings," *Ophthalmology*, Vol. 85, pp. 82-105.

**DRSRG** [1979] "Four Risk Factors for Severe Visual Loss in Diabetic Retinopathy," *Archives of Ophthalmology*, Vol. 97, pp. 654-655.

**DRSRG** [1980] "Photocoagulation Treatment of Proliferative Diabetic Retinopathy: A Short Report of Long-Range Results," Diabetic Retinopathy Study Report Number 4, Amsterdam: *Excerpta Medica*, 789-794.

**DRSRG** [1981] "Photocoagulation Treatment of Proliferative Diabetic Retinopathy: Relationship of Adverse Treatment Effects to Retinopathy Severity," *Developmental Ophthalmology*, Vol. 2, pp. 248-261.

**DRSRG** [1981] "Report 6: Design, Methods and Baseline Results," *Investigative Ophthalmology and Visual Science*, Vol. 21/1, Part 2, pp. 149-209.

**DRSRG** [1981] "Report 7: A Modification of the Airlie House Classification of Diabetic Retinopathy," *Investigative Ophthalmology and Visual Science*, Vol. 21/1, Part 2, pp. 210-226.

**DRSRG** [1981] "Photocoagulation Treatment of Proliferative Diabetic Retinopathy: Clinical Application of DRS Findings," *Ophthalmology*, Vol. 88, No. 7, pp. 583-600.

**DRSRG** [1984] "Assessing Possible Late Treatment Effects in Stopping a Clinical Trial Early: A Case Study," *Controlled Clinical Trials*, Vol. 5, pp. 373-381.

The Diabetic Retinopathy Study (DRS) begun in 1971 under sponsorship of the National Eye Institute, USA. Randomised controlled trial to examine effects of photocoagulation therapy. Preliminary Report stated that: "The assessment of treatment effects in this report is limited to visual acuity and visual field." No references in subsequent reports to any benefits of treatment other than visual acuity/visual field, e.g., Report 9: "A five-year follow-up was planned for each patient, and the principal response for gauging the efficacy of the treatment was the occurrence of severe visual loss ("blindness"). This was defined as visual acuity less than 5/200 at two or more consecutive follow-up visits scheduled at 4-month intervals."

**Dobree, J.H. and Taylor, E.** [1973] "Treatment of Proliferative Diabetic Retinopathy by Repeated Light Coagulation: A 7-Year Review," *British Journal of Ophthalmology*, 57, pp. 73-80.

"In 1968 the results of treating 42 eyes from 35 patients suffering from proliferative diabetic retinopathy were presented at the Airlie House Symposium on Diabetic Retinopathy. We have now treated 174 eyes from 126 consecutive cases and present the results of the series which has now been running for over 7 years." In comparing treated cases

with controls, "the parameter of the retention of the original visual acuity" was used to assess changes over time.

## E

**Early Treatment Diabetic Retinopathy Study Research Group** [1985] "Photocoagulation for Diabetic Macular Edema," *Archives of Ophthalmology*, Vol. 103, pp. 1796-1806.

National Eye Institute-supported multi-centre R.C.T. designed to evaluate, inter alia, photocoagulation treatment for patients with non-proliferative or early proliferative diabetic retinopathy.

Outcome assessment: "Visual acuity, the primary means of gauging treatment effects in this study, was measured at each visit by an examiner who did not know the treatment assignment (masked)..."

Other visual function tests were performed, e.g., calculation of visual field and colour vision scores. Authors also present results based on patients' impressions of change in visual function:

"In 16 of the 23 clinical centres.....the results for improvement were statistically significant (Z>2.57)."

Study includes a "Subjective Evaluation of Visual Functions" questionnaire, containing:

Part I. Identifying Information

Part II. Current Status (series of questions about vision in both eyes).

Part III. Comparison (between vision today and vision one year ago).

Part IV. Administrative Matters Visual Acuity Supplemental Questions.

## G

**Genensky, S.M.** [1976] "Acuity Measurements—Do They Indicate How Well a Partially Sighted Person Functions or Could Function?" *American Journal Optometry and Physiological Optics*, Vol. 53, No. 12, pp. 809-812.

Visual acuity measurement, it is argued, does not tell us very much about what a patient has accomplished or can be expected to accomplish with his residual vision. For example, it does not necessarily say anything about the quality of the image perceived by the patient.

Measuring both the near and distance visual acuities is important because the two measurements give a much clearer picture of how well a partially sighted subject sees. "In addition to these measurements, I believe in having knowledge of at least the subject's gross visual fields. By "gross visual fields," I mean information concerning whether or not the patient has significant scotomas in one or more quadrants of his fields or in his central fields. With these data, I can better judge how well the subject sees the world around him . . . and I can explain to the subject our conception of his visual abilities."

The author's main complaint with distance VA measurements is their current use as "the major criterion by which to judge whether a person is blind or sighted." Indeed he adds that "the definition of legal blindness in this country, based so heavily on distant visual acuity, has done more harm to partially sighted people than any other definition used by our federal and state governments."

**Genensky, S.M., Berry, S.H., Bikson, T.H.** et al. [1979] Visual Environmental Adaptation Problems of the Partially Sighted, Santa Monica, California, Centre for the Partially Sighted.

"Questionnaire Codebook and Mobility Rating Form" (by Berry and Genensky) serves as basic documentation of data collected during interviews with the 94 participants, and as a reference for the Final Report. It is also "expected to serve as a guide for future users of data collected as part of this project and for designers of new questionnaires for partially sighted populations."

Extremely detailed codebook which provides for each question:

(a) the wording of the question;

(b) changes in question wording and coding made during data collection or reduction;

(c) frequencies for open-ended responses;

(d) code categories developed for most open-ended questions; and

(e) frequencies for open-ended questions.

## H

**Haase, K.W. and Bryant, E.E.** [1973] "Development of a Scale Designed to Measure Functional Distance Vision Loss Using an Interview Technique," *Proceedings of The American Statistical Association*, Social Statistics Section, pp. 274-279.

Study aim was to develop and test three scales designed to measure functional vision loss by use of an interview technique. Paper here presents preliminary assessment of a distance vision scale when used alone and when used in conjunction with a self-evaluation scale. Distance scale consisted of five questions ordered in the form of a Guttman Scale. Authors were "generally encouraged by the distance scales" ability to classify populations according to perceived functional vision loss."

Distance Scale Questions:

1. (When wearing glasses) can you see well enough to recognise a friend if you get close to his face?
2. (When wearing glasses) can you see well enough to recognise a friend who is an arm's length away?
3. (When wearing glasses) can you see well enough to recognise a friend across a room?
4. (When wearing glasses) can you see well enough to recognise a friend across a street?
5. Do you have any problems seeing distant objects?

**Hiller, R. and Kruger, D.E.** [1983] "Validity of a Survey Question to Study Diabetic Retinopathy," *American Journal of Public Health*, Vol. 73, No. 1, pp. 93-95.

Paper presents results of a comparison of questionnaire response on "trouble seeing" and visual acuity measurement. Provides estimates of the sensitivity and specificity of

the question "Do (did) you have trouble with your vision even when wearing glasses or contact lenses?"

For each age group, authors found the sensitivity to be low, indicating that this question failed to identify most adults who have impaired visual acuity. The specificity, however, was high, suggesting that the question successfully identified subjects without visual acuity impairment.

Suggests that: "the single questions used in surveys to date have been inadequate for many purposes. Perhaps a series of task-related questions . . . will yield a higher sensitivity as well as a more accurate reflection of the marked increase in impairment with increasing age."

## J/K

**Jastrzembska, Z.S.** [1973] "Social and Psychological Aspects of Blindness: A Sampling of the Literature," *American Foundation for the Blind Research Bulletin*, No. 25, pp. 169-173.

Paper emphasises the vast scope of socio-psychological aspects of blindness and notes that the various aspects discussed in the literature do form an "integrated whole." Attempts to provide an overview of this emerging area using selected references.

For example, the literature points to the intense anger and frustration which the limitations imposed by lack of vision can produce even in the very young congenitally blind. However, when blindness occurs in later life, the difficulties it creates are less fundamental. The problem then is to learn new ways to continue already-established activities, and to accept the curtailment of some.

**Keltner, J.L.** et al. [1985] "The California Syndrome: Functional Visual Complaints with Potential Economic Impact," *Ophthalmology*, Vol. 92, No. 3, pp. 427-435.

The records of 84 patients with functional vision loss (59 adults, 25 children) were reviewed over a 6-year period. In the young patients, psychosocial problems relating to parental divorce, poor school performance and attention-getting behaviour were common, while in adult patients, financial gain was present in 86% of individuals. Over 90% of the claimed abnormalities related to the afferent visual system; 53% of patients had underlying organic problems with functional overlay. Inconsistent test results were the most reliable means of identifying the malingering patients. In California, both ophthalmologists and optometrists are able to declare a patient legally blind as well as visually disabled for the purpose of income tax, workers' compensation and disability benefits. The potential financial impact of malingering patients is discussed.

## M/O

**Macular Photocoagulation Study Group** [1986] "Argon Laser Photocoagulation for Neovascular Maculopathy: Three Year Results from Randomised Controlled Trials," *Archives of Ophthalmology*, Vol. 104, pp. 694-701.

The M.P.S.G. is conducting three multicentre R.C.T.'s designed to determine the effectiveness of argon laser photocoagulation. After 3 years, "The beneficial effects of argon laser treatment have persisted in all three studies, in spite of the fact that some eyes in the "no-treatment" groups were treated later in their clinical course as a result of the evidence of early benefit."

Statistical Methods: "Differences in visual acuity between eyes assigned to the argon laser photocoagulation treatment group and to the no-treatment group were analysed using change in visual acuity from baseline as the primary outcome variable."

Useful article in setting out the procedures used to calculate the confidence interval on the relative risk of loss of visual acuity, and to compare distributions of individual visual acuity scored and of changes in scores. Also involved use of survival analysis and product-limit methods (to account for recovery of vision) to evaluate differences during follow-up time.

**Oehler-Giarratana, J.** [1978] "Meeting the Psychosocial and Rehabilitative Needs of the Visually-Impaired Diabetic," *Journal of Visual Impairment and Blindness*, Vol. 72, Part 9, pp. 358-361.

Classifies the psychosocial and visual problems caused by proliferating diabetic retinopathy into four stages:

(1) discovery of diagnosis;

(2) developing awareness of diagnosis;

(3) realisation of disability;

(4) stabilisation of vision.

**Oehler-Giarratana, J.** [1980] An Exploratory Study of Psychological Reactions to Visual Loss and Blindness in Patients with Diabetic Retinopathy, Ph.D. Thesis, Boston, MA, Boston University.

Cross-sectional assessment of 33 adults with diabetic retinopathy or maculopathy and recent visual loss of <5 years. (VA of 20/40 or worse in the better eye). Used a semi-structured interview and two standardised self-report scales for depression [Zung's Self-Rating Depression Scale and the MMPI Depression Scale] to measure psychological responses to visual loss. In the interval since visual loss, 33% of the sample reported severe depression," 69% reported work difficulties related to vision, 39% lost jobs, 73% could no longer drive, and 39% started receiving financial support from social security income.

## T/W

**Townsend, C., Bailey, J and Kohner, E.M.** [1979] "Xenon-arc Photocoagulation in the Treatment of Diabetic Maculopathy," *Transactions of the Ophthalmological Society of the UK*, 99, pp. 13-16.

Represents on behalf of the British Multicentre Photocoagulation Group, the effects over a period of up to five years on the visual acuity of patients with diabetic maculopathy of treatment with xenon-arc photocoagulation. [see also British Multicentre Photocoagulation Trial, 1975]

R.C.T. where one eye of each patient included in the trial was randomly allocated for treatment, the other eye being left as an untreated control. Conclusion: diabetic maculopathy is treatable by xenon-arc photocoagulation provided that the initial visual acuity is better than 6/36. In all such cases the difference in deterioration of visual acuity was

greater in the untreated than in the treated eyes, i.e., changes in visual acuity were the yardstick for measuring benefits of treatment.

**Westat, Inc.** [1976] Summary and Critique of Available Data on the Prevalence and Economic and Social Costs of Visual Disorders and Disabilities, Unpublished, Rockville, Maryland, Feb. 16.

Summary of findings of survey and analysis of available information concerning the extent and costs of eye disorders and blindness in the USA. Study conducted for National Eye Institute under contract by Westat, Inc.

" . . . among the chronic diseases that restrict the ability of Americans to lead normal productive lives, visual impairment ranks third after heart disease and arthritis. Thus, the chief impact of eye diseases and eye accidents is not mortality, but physical limitation and financial burden."

**Wulsin, L.R., Jacobson, A.M. and Rand, L.I.** [1987] "Psychosocial Aspects of Diabetic Retinopathy," *Diabetes Care*, Vol. 10, No. 3, pp. 367- 373.

Diabetic retinopathy, particularly in the more advanced stages, poses many difficult psychosocial problems and demands major adjustment by the patient. Our review of this literature has identified specific problems relevant to patient care, future research, and public policy. For example, proliferative retinopathy often leads to at least partial visual impairment, psychiatric symptoms, and difficulties with glycaemic control. Partial visual impairment appears to cause as much psychosocial disruption as severe blindness. This suggests that most rehabilitation programmes that serve the legally blind may come too late in the course of this illness. This review emphasises the paucity of past research on psychosocial aspects of diabetic retinopathy and raises some questions for future research.

HV1714 Drummond, Michael F.
D844 Measuring the quality of
c.2 life of people with
visual impairment:
proceedings of a workshop.

DATE DUE

| | | | |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

DEMCO

NIH Publication No. 90-3078